

Quantitative Techniques for Management

**MBA First Year
Paper No. 6**

**School of Distance Education
Bharathiar University, Coimbatore - 641 046**

Author: P N Mishra & S Jaisankar

Copyright © 2007, Bharathiar University
All Rights Reserved

Edited by: Dr. Subodh Kesharwani
Produced and Printed
by
EXCEL BOOKS PRIVATE LIMITED
A-45, Naraina, Phase-I,
New Delhi-110028
for
SCHOOL OF DISTANCE EDUCATION
Bharathiar University
Coimbatore-641046

CONTENTS

	Page No.
Unit -I	
Lesson 1	Quantitative Techniques – Introduction
Lesson 2	Measures of Central Tendency
Lesson 3	Mathematical Model
Lesson 4	Linear Programming: Graphical Method
Lesson 5	Linear Programming: Simplex Method
Unit -II	
Lesson 6	Transportation Model
Lesson 7	Assignment Model
Unit -III	
Lesson 8	Network Model
Lesson 9	Waiting Model (Queuing Theory)
Unit -IV	
Lesson 10	Probability
Lesson 11	Theoretical Probability Distributions
Lesson 12	Probability Distribution of a Random Variable
Unit -V	
Lesson 13	Inventory Model
Lesson 14	Game Theory
Lesson 15	Simulation

QUANTITATIVE TECHNIQUES FOR MANAGEMENT

Number of Credit Hours : 3 (Three)

Subject Description: This course presents the various mathematical models, networking, probability, inventory models and simulations for managerial decisions.

Goals: To enable the students to learn techniques of operations research and resources management and their application in decision making in the management.

Objectives: On successful completion of the course the students should have:

1. Understood the basic of the quantitative techniques.
2. Learnt the feasible solution and optimum solution for the resource management.
3. Learnt the time estimation and critical path for project.
4. Learnt about the application of probability techniques in the decision making.
5. Learnt the various inventory models and simulations in the resource planning and management.

UNIT I

QT – Introduction – Measures of Central Tendency – Mean, Median, Mode.

Mathematical Models – deterministic and probabilistic – simple business examples – OR and optimization models – Linear Programming – formulation – graphical solution –simplex – solution.

UNIT II

Transportation model – Initial Basic Feasible solutions – optimum solution for non – degeneracy and degeneracy model – Trans-shipment Model – Assignment Model – Travelling Salesmen problem.

UNIT III

Network Model – networking – CPM – critical path – Time estimates – critical path – crashing, Resource levelling, Resources planning. Waiting Line Model – Structure of model – M/M/1 for infinite population.

UNIT IV

Probability – definitions – addition and multiplication Rules (only statements) – simple business application problems – probability distribution – expected value concept – theoretical probability distributions – Binomial, Poisson and Normal – Simple problems applied to business.

UNIT V

Inventory Models – Deterministic – EOQ – EOQ with Price Breaks – Probabilistic Inventory Models - Probabilistic EOQ model – Game theory-zero sum games: Arithmetic and Graphical Method.

Simulation – types of simulation – Monte Carlo simulation – simulation problems.

Decision Theory – Pay off tables – decision criteria – decision trees.

Unit-I

LESSON

1

QUANTITATIVE TECHNIQUES – INTRODUCTION

CONTENTS

- 1.0 Aims and Objectives
- 1.1 Introduction
- 1.2 Historical Development
- 1.3 About Quantitative Technique
- 1.4 Methodology of Quantitative Techniques
 - 1.4.1 Formulating the Problem
 - 1.4.2 Defining the Decision Variables and Constraints
 - 1.4.3 Developing a Suitable Model
 - 1.4.4 Acquiring the Input Data
 - 1.4.5 Solving the Model
 - 1.4.6 Validating the Model
 - 1.4.7 Implementing the Results
- 1.5 Advantages of Mathematical Modelling
- 1.6 Scope of Quantitative Technique
- 1.7 Statistics : An Introduction
 - 1.7.1 Origin and Growth of Statistics
 - 1.7.2 Meaning and Definition of Statistics
 - 1.7.3 Statistics as Data
 - 1.7.4 Statistics as a Science
 - 1.7.5 Statistics as a Science different from Natural Sciences
 - 1.7.6 Statistics as a Scientific Method
 - 1.7.7 Statistics as a Science or an Art
- 1.8 Let us Sum Up
- 1.9 Lesson-end Activities
- 1.10 Keywords
- 1.11 Questions for Discussion
- 1.12 Terminal Questions
- 1.13 Model Answers to Questions for Discussion
- 1.14 Suggested Readings

1.0 AIMS AND OBJECTIVES

In this first lesson we discuss the distinguished approaches to quantitative techniques and its various applications in management, statistical analysis and other industries. Here we will discuss the approaches of quantitative techniques.

1.1 INTRODUCTION

Scientific methods have been man's outstanding asset to pursue an ample number of activities. It is analysed that whenever some national crisis, emerges due to the impact of political, social, economic or cultural factors the talents from all walks of life amalgamate together to overcome the situation and rectify the problem. In this chapter we will see how the quantitative techniques had facilitated the organization in solving complex problems on time with greater accuracy. The historical development will facilitate in managerial decision-making & resource allocation, The methodology helps us in studying the scientific methods with respect to phenomenon connected with human behaviour like formulating the problem, defining decision variable and constraints, developing a suitable model, acquiring the input data, solving the model, validating the model, implementing the results. The major advantage of mathematical model is that its facilitates in taking decision faster and more accurately.

Managerial activities have become complex and it is necessary to make right decisions to avoid heavy losses. Whether it is a manufacturing unit, or a service organization, the resources have to be utilized to its maximum in an efficient manner. The future is clouded with uncertainty and fast changing, and decision-making – a crucial activity – cannot be made on a trial-and-error basis or by using a thumb rule approach. In such situations, there is a greater need for applying scientific methods to decision-making to increase the probability of coming up with good decisions. Quantitative Technique is a scientific approach to managerial decision-making. The successful use of Quantitative Technique for management would help the organization in solving complex problems on time, with greater accuracy and in the most economical way. Today, several scientific management techniques are available to solve managerial problems and use of these techniques helps managers become explicit about their objectives and provides additional information to select an optimal decision. This study material is presented with variety of these techniques with real life problem areas.

1.2 HISTORICAL DEVELOPMENT

During the early nineteen hundreds, Fredrick W. Taylor developed the scientific management principle which was the base towards the study of managerial problems. Later, during World War II, many scientific and quantitative techniques were developed to assist in military operations. As the new developments in these techniques were found successful, they were later adopted by the industrial sector in managerial decision-making and resource allocation. The usefulness of the Quantitative Technique was evidenced by a steep growth in the application of scientific management in decision-making in various fields of engineering and management. At present, in any organization, whether a manufacturing concern or service industry, Quantitative Techniques and analysis are used by managers in making decisions scientifically.

Check Your Progress 1.1

Explain with the help of example some of the important Quantitative Techniques used in modern business and in industrial unit.

Contd....

- Notes:**
- (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-

1.3 ABOUT QUANTITATIVE TECHNIQUE

Quantitative Techniques adopt a scientific approach to decision-making. In this approach, past data is used in determining decisions that would prove most valuable in the future. The use of past data in a systematic manner and constructing it into a suitable model for future use comprises a major part of scientific management. For example, consider a person investing in fixed deposit in a bank, or in shares of a company, or mutual funds, or in Life Insurance Corporation. The expected return on investments will vary depending upon the interest and time period. We can use the scientific management analysis to find out how much the investments made will be worth in the future. There are many scientific method software packages that have been developed to determine and analyze the problems.

In case of complete non-availability of past data, quantitative factors are considered in decision-making. In cases where the scope of quantitative data is limited, qualitative factors play a major role in making decisions. Qualitative factors are important situations like sudden change in tax-structures, or the introduction of breakthrough technologies. Application of scientific management and Analysis is more appropriate when there is not much of variation in problems due to external factors, and where input values are steady. In such cases, a model can be developed to suit the problem which helps us to take decisions faster. In today's complex and competitive global marketplace, use of Quantitative Techniques with support of qualitative factors is necessary.

Quantitative Technique is the scientific way to managerial decision-making, while emotion and guess work are not part of the scientific management approach. This approach starts with data. Like raw material for a factory, this data is manipulated or processed into information that is valuable to people making decision. This processing and manipulating of raw data into meaningful information is the heart of scientific management analysis.

Check Your Progress 1.2

Do you think the day will come when all decision in a business unit are made with assistance of quantitative techniques? Give reasons for your answer.

- Notes:**
- (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.

Contd....

- (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

1.4 METHODOLOGY OF QUANTITATIVE TECHNIQUES

The methodology adopted in solving problems is as follows:

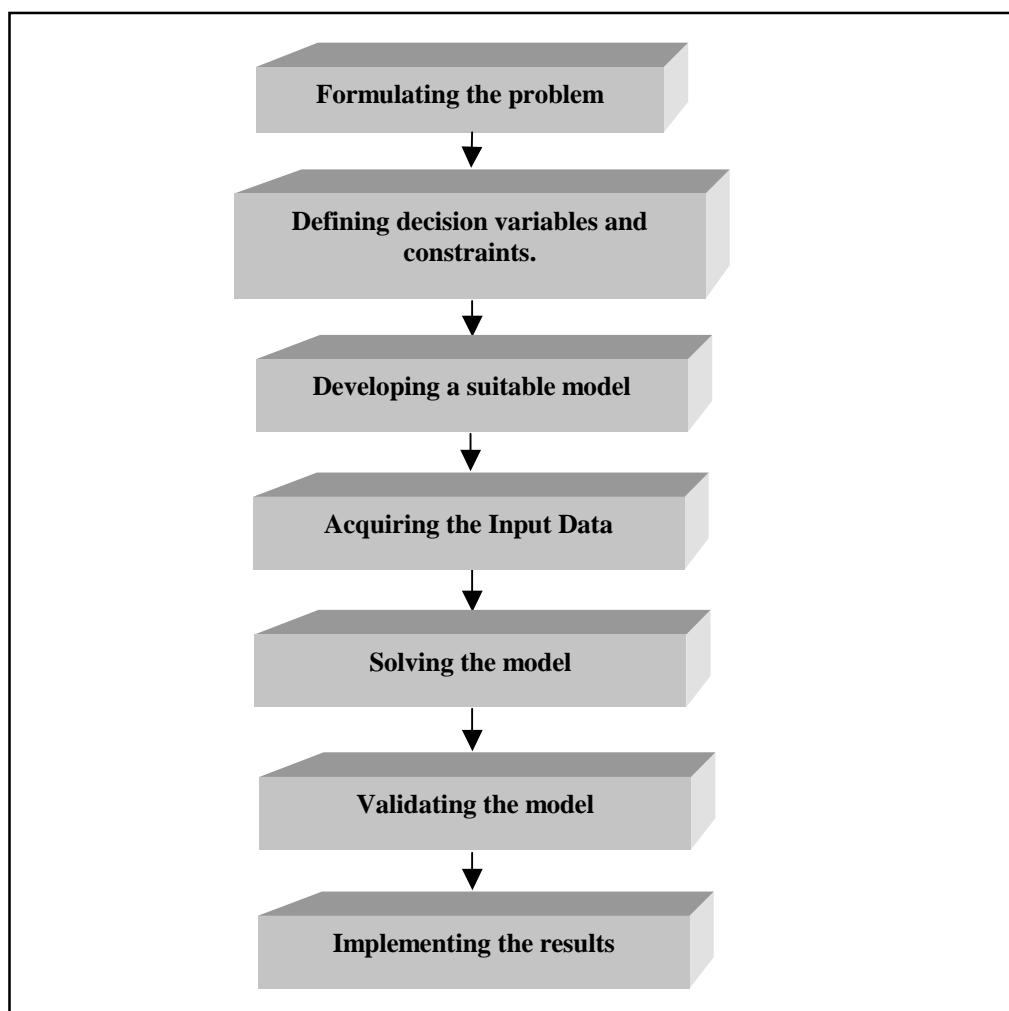


Figure 1.1

1.4.1 Formulating the Problem

As a first step, it is necessary to clearly understand the problem situations. It is important to know how it is characterized and what is required to be determined. Firstly, the key decision and the objective of the problem must be identified from the problem. Then, the number of decision variables and the relationship between variables must be determined. The measurable guarantees that are represented through these variables are notified. The practical limitations or constraints are also inferred from the problem.

1.4.2 Defining the Decision Variables and Constraints

In a given problem situation, defining the key decision variables are important. Identifying these variables helps us to develop the model. For example, consider a manufacturer who is manufacturing three products A, B and C using two machines, I and II. Each unit of product A takes 2 minutes on machine I and 5 minutes on machine II. Product B takes 1 minute on machine I and 3 minutes on machine II. Similarly, product C takes 4 minutes and 6 minutes on machine I and machine II, respectively. The total available time on machine I and machine II are 100 hours and 120 hours, respectively. Each unit of A yields a profit of Rs. 3.00, B yields Rs. 4.00 and C yields Rs. 5.00. What should be level of production of products A, B and C that should be manufactured by the company so as to maximize the profit?

The decision variables, objective and constraints are identified from the problem.

The company is manufacturing three products A, B and C. Let A be x_1 , B be x_2 and C be x_3 . x_1 , x_2 and x_3 are the three decision variables in the problem. The objective is to maximize the profits. Therefore, the problem is to maximize the profit, i.e., to know how many units of x_1 , x_2 and x_3 are to be manufactured. There are two machines available, machine I and machine II with total machine hours available as 100 hours and 120 hours. The machine hours are the resource constraints, i.e., the machines cannot be used more than the given number of hours.

To summarize,

- Key decision : How many units of x_1 , x_2 and x_3 are to be manufactured
- Decision variables : x_1 , x_2 and x_3
- Objective : To maximize profit
- Constraint : Machine hours

1.4.3 Developing a Suitable Model

A model is a mathematical representation of a problem situation. The mathematical model is in the form of expressions and equations that replicate the problem. For example, the total profit from a given number of products sold can be determined by subtracting selling price and cost price and multiplying the number of units sold. Assuming selling price, sp as Rs. 40 and cost price, cp as Rs. 20, the following mathematical model expresses the total profit, tp earned by selling number of unit x.

$$TP = (SP - CP) x$$

$$= (40 - 20) x$$

$$TP = 20 x$$

Now, this mathematical model enables us to identify the real situation by understanding the model. The models can be used to maximize the profits or to minimize the costs. The applications of models are wide, such as:

- Linear Programming Model
- Integer Programming
- Sensitivity Analysis
- Goal Programming
- Dynamic Programming
- Non Linear Programming
- Queuing Theory
- Inventory Management Techniques

- PERT/CPM (Network Analysis)
- Decision Theory
- Games Theory
- Transportation and Assignment Models.

1.4.4 Acquiring the Input Data

Accurate data for input values are essential. Even though the model is well constructed, it is important that the input data is correct to get accurate results. Inaccurate data will lead to wrong decisions.

1.4.5 Solving the Model

Solving is trying for the best result by manipulating the model to the problem. This is done by checking every equation and its diverse courses of action. A trial and error method can be used to solve the model that enables us to find good solutions to the problem.

1.4.6 Validating the Model

A validation is a complete test of the model to confirm that it provides an accurate representation of the real problem. This helps us in determining how good and realistic the solution is. During the model validation process, inaccuracies can be rectified by taking corrective actions, until the model is found to be fit.

1.4.7 Implementing the Results

Once the model is tested and validated, it is ready for implementation. Implementation involves translation/application of solution in the company. Close administration and monitoring is required after the solution is implemented, in order to address any proposed changes that call for modification, under actual working conditions.

1.5 ADVANTAGES OF MATHEMATICAL MODELLING

The advantages of mathematical modelling are many:

- (a) Models exactly represent the real problem situations.
- (b) Models help managers to take decisions faster and more accurately.
- (c) Models save valuable resources like money and time.
- (d) Large and complex problems can be solved with ease.
- (e) Models act as communicators to others by providing information and impact in changing conditions.

Check Your Progress 1.3

“Quantitative Technique is a very powerful tools and analytical process that offers the presentation of an optimum solutions in spite of its limitations”. Discuss.

- Notes:**
- (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

1.6 SCOPE OF QUANTITATIVE TECHNIQUE

The scope and areas of application of scientific management are very wide in engineering and management studies. Today, there are a number of quantitative software packages available to solve the problems using computers. This helps the analysts and researchers to take accurate and timely decisions. This book is brought out with computer based problem solving. A few specific areas are mentioned below.

- ***Finance and Accounting:*** Cash flow analysis, Capital budgeting, Dividend and Portfolio management, Financial planning.
- ***Marketing Management:*** Selection of product mix, Sales resources allocation and Assignments.
- ***Production Management:*** Facilities planning, Manufacturing, Aggregate planning, Inventory control, Quality control, Work scheduling, Job sequencing, Maintenance and Project planning and scheduling.
- ***Personnel Management:*** Manpower planning, Resource allocation, Staffing, Scheduling of training programmes.
- ***General Management:*** Decision Support System and Management of Information Systems, MIS, Organizational design and control, Software Process Management and Knowledge Management.

From the various definitions of Quantitative Technique it is clear that scientific management has got wide scope. In general, whenever there is any problem simple or complicated the scientific management technique can be applied to find the best solutions. In this head we shall try to find the scope of M.S. by seeing its application in various fields of everyday life this include define operation too.

Check Your Progress 1.4

Discuss the significance and scope of Quantitative Techniques in modern business management.

- Notes:**
- (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-
-

1.7 STATISTICS : AN INTRODUCTION

1.7.1 Origin and Growth of Statistics

Statistics, as a subject, has a very long history. The origin of STATISTICS is indicated by the word itself which seems to have been derived either from the Latin word 'STATUS' or from the Italian word 'STATISTA' or may be from the German word 'STATISTIK.' The meaning of all these words is 'political state'. Every State administration in the past collected and analysed data. The data regarding population gave an idea about the possible military strength and the data regarding material wealth of a country gave an idea about the possible source of finance to the State. Similarly, data were collected for other purposes also. On examining the historical records of various ancient countries, one might find that almost all the countries had a system of collection of data. In ancient Egypt, the data on population and material wealth of the country were collected as early as 3050 B.C., for the construction of pyramids. Census was conducted in Jidda in 2030 B.C. and the population was estimated to be 38,00,000. The first census of Rome was done as early as 435 B.C. After the 15th century the work of publishing the statistical data was also started but the first analysis of data on scientific basis was done by Captain John Graunt in the 17th century. His first work on social statistics, 'Observation on London Bills of Mortality' was published in 1662. During the same period the gamblers of western countries had started using statistics, because they wanted to know the more precise estimates of odds at the gambling table. This led to the development of the 'Theory of Probability'.

Ancient India also had the tradition of collection of statistical data. In ancient works, such as Manusmriti, Shukraniti, etc., we find evidences of collection of data for the purpose of running the affairs of the State where population, military force and other resources have been expressed in the form of figures. The fact and figures of the Chandragupta Mauraya's regime are described in 'Kautilya's Arthashastra'. Statistics were also in use during the Mughal period. The data were collected regarding population, military strength, revenue, land revenue, measurements of land, etc. The system of data collection was described in Tuzuk - i - Babri and Ain-i-Akbari. During Akbar's period, his revenue minister, Raja Todarmal, made a well organised survey of land for the collection of land revenue. During the British period too, statistics were used in various areas of activities.

Although the tradition of collection of data and its use for various purposes is very old, the development of modern statistics as a subject is of recent origin. The development of the subject took place mainly after sixteenth century. The notable mathematicians who contributed to the development of statistics are Galileo, Pascal, De-Mere, Farment and Cardeno of the 17th century. Then in later years the subject was developed by Abraham De Moivre (1667 - 1754), Marquis De Laplace (1749 - 1827), Karl Friedrich Gauss (1777 - 1855), Adolphe Quetelet (1796 - 1874), Francis Galton (1822 - 1911), etc. Karl Pearson (1857 - 1937), who is regarded as the father of modern statistics, was greatly motivated by the researches of Galton and was the first person to be appointed as Galton Professor in the University of London. William S. Gosset (1876 - 1937), a student of Karl Pearson, propounded a number of statistical formulae under the pen-name of 'Student'. R.A. Fisher is yet another notable contributor to the field of statistics. His book 'Statistical Methods for Research Workers', published in 1925, marks the beginning of the theory of modern statistics.

The science of statistics also received contributions from notable economists such as Augustin Cournot (1801 - 1877), Leon Walras (1834 - 1910), Vilfredo Pareto (1848 - 1923), Alfred Marshall (1842 - 1924), Edgeworth, A.L. Bowley, etc. They gave an applied form to the subject.

Among the noteworthy Indian scholars who contributed to statistics are P.C. Mahalanobis, V.K.R.V. Rao, R.C. Desai, P.V. Sukhatme, etc.

1.7.2 Meaning and Definition of Statistics

The meaning of the word 'Statistics' is implied by the pattern of development of the subject. Since the subject originated with the collection of data and then, in later years, the techniques of analysis and interpretation were developed, the word 'statistics' has been used in both the plural and the singular sense. Statistics, in plural sense, means a set of numerical figures or data. In the singular sense, it represents a method of study and therefore, refers to statistical principles and methods developed for analysis and interpretation of data.

Statistics has been defined in different ways by different authors. These definitions can be broadly classified into two categories. In the first category are those definitions which lay emphasis on statistics as data whereas the definitions in second category emphasise statistics as a scientific method.

1.7.3 Statistics as Data

Statistics used in the plural sense implies a set of numerical figures collected with reference to a certain problem under investigation. It may be noted here that any set of numerical figures cannot be regarded as statistics. There are certain characteristics which must be satisfied by a given set of numerical figures in order that they may be termed as statistics. Before giving these characteristics it will be advantageous to go through the definitions of statistics in the plural sense, given by noted scholars.

1. "Statistics are numerical facts in any department of enquiry placed in relation to each other." - A.L. Bowley

The main features of the above definition are :

- (i) Statistics (or Data) implies numerical facts.
- (ii) Numerical facts or figures are related to some enquiry or investigation.
- (iii) Numerical facts should be capable of being arranged in relation to each other.

On the basis of the above features we can say that data are those numerical facts which have been expressed as a set of numerical figures related to each other and to some area of enquiry or research. We may, however, note here that all the characteristics of data are not covered by the above definition.

2. "By statistics we mean quantitative data affected to a marked extent by multiplicity of causes." - Yule & Kendall

This definition covers two aspects, i.e., the data are quantitative and affected by a large number of causes.

3. "Statistics are classified facts respecting the conditions of the people in a state—especially those facts which can be stated in numbers or in tables of numbers or in any other tabular or classified arrangement." - Webster

4. "A collection of noteworthy facts concerning state, both historical and descriptive." - Achenwall

Definitions 3 and 4, given above, are not comprehensive because these confine the scope of statistics only to facts and figures related to the conditions of the people in a state. However, as we know that data are now collected on almost all the aspects of human and natural activities, it cannot be regarded as a state-craft only.

5. "Statistics are measurements, enumerations or estimates of natural or social phenomena, systematically arranged, so as to exhibit their interrelations."

- L.R. Connor

This definition also covers only some but not all characteristics of data.

6. "By statistics we mean aggregate of facts affected to a marked extent by a multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other." - **H. Secrist**

This definition can be taken as a comprehensive definition of statistics since most of the characteristics of statistics are covered by it.

Characteristics of Statistics as Data

On the basis of the above definitions we can now state the following characteristics of statistics as data :

1. **Statistics are numerical facts:** In order that any set of facts can be called as statistics or data, it must be capable of being represented numerically or quantitatively. Ordinarily, the facts can be classified into two categories : (a) Facts that are measurable and can be represented by numerical measurements. Measurement of heights of students in a college, income of persons in a locality, yield of wheat per acre in a certain district, etc., are examples of measurable facts. (b) Facts that are not measurable but we can feel the presence or absence of the characteristics. Honesty, colour of hair or eyes, beauty, intelligence, smoking habit etc., are examples of immeasurable facts. Statistics or data can be obtained in such cases also, by counting the number of individuals in different categories. For example, the population of a country can be divided into three categories on the basis of complexion of the people such as white, whitish or black.

2. **Statistics are aggregate of facts:** A single numerical figure cannot be regarded as statistics. Similarly, a set of unconnected numerical figures cannot be termed as statistics. Statistics means an aggregate or a set of numerical figures which are related to one another. The number of cars sold in a particular year cannot be regarded as statistics. On the other hand, the figures of the number of cars sold in various years of the last decade is statistics because it is an aggregate of related figures. These figures can be compared and we can know whether the sale of cars has increased, decreased or remained constant during the last decade.

It should also be noted here that different figures are comparable only if they are expressed in same units and represent the same characteristics under different situations. In the above example, if we have the number of Ambassador cars sold in 1981 and the number of Fiat cars sold in 1982, etc., then it cannot be regarded as statistics. Similarly, the figures of, say, measurement of weight of students should be expressed in the same units in order that these figures are comparable with one another.

3. **Statistics are affected to a marked extent by a multiplicity of factors:** Statistical data refer to measurement of facts in a complex situation, e.g., business or economic phenomena are very complex in the sense that there are a large number of factors operating simultaneously at a given point of time. Most of these factors are even difficult to identify. We know that quantity demanded of a commodity, in a given period, depends upon its price, income of the consumer, prices of other commodities, taste and habits of the consumer. It may be mentioned here that these factors are only the main factors but not the only factors affecting the demand of a commodity. Similarly, the sale of a firm in a given period is affected by a large number of factors. Data collected under such conditions are called statistics or statistical data.

4. **Statistics are either enumerated or estimated with reasonable standard of accuracy:** This characteristic is related to the collection of data. Data are collected either by counting or by measurement of units or individuals. For example, the number of smokers in a village are counted while height of soldiers is measured.

We may note here that if the area of investigation is large or the cost of measurement is high, the statistics may also be collected by examining only a fraction of the total area of investigation.

When statistics are being obtained by measurement of units, it is necessary to maintain a reasonable degree or standard of accuracy in measurements. The degree of accuracy needed in an investigation depends upon its nature and objectivity on the one hand and upon time and resources on the other. For example, in weighing of gold, even milligrams may be significant whereas, for weighing wheat, a few grams may not make much difference. Sometimes, a higher degree of accuracy is needed in order that the problem, to be investigated, gets highlighted by the data. Suppose the diameter of bolts produced by a machine are measured as 1.546 cms, 1.549 cms, 1.548 cms, etc. If, instead, we obtain measurements only up to two places after decimal, all the measurements would be equal and as such nothing could be inferred about the working of the machine. In addition to this, the degree of accuracy also depends upon the availability of time and resources. For any investigation, a greater degree of accuracy can be achieved by devoting more time or resources or both. As will be discussed later, in statistics, generalisations about a large group (known as population) are often made on the basis of small group (known as sample). It is possible to achieve this by maintaining a reasonable degree of accuracy of measurements. Therefore, it is not necessary to always have a high degree of accuracy but whatever degree of accuracy is once decided must be uniformly maintained throughout the investigation.

5. ***Statistics are collected in a systematic manner and for a predetermined purpose:*** In order that the results obtained from statistics are free from errors, it is necessary that these should be collected in a systematic manner. Haphazardly collected figures are not desirable as they may lead to wrong conclusions. Moreover, statistics should be collected for a well defined and specific objective, otherwise it might happen that the unnecessary statistics are collected while the necessary statistics are left out. Hence, a given set of numerical figures cannot be termed as statistics if it has been collected in a haphazard manner and without proper specification of the objective.
6. ***Statistics should be capable of being placed in relation to each other:*** This characteristic requires that the collected statistics should be comparable with reference to time or place or any other condition. In order that statistics are comparable it is essential that they are homogeneous and pertain to the same investigation. This can be achieved by collecting data in identical manner for different periods or for different places or for different conditions.

Hence, any set of numerical facts possessing the above mentioned characteristics can be termed as statistics or data.

Example 1: Would you regard the following information as statistics? Explain by giving reasons.

- (i) The height of a person is 160 cms.
- (ii) The height of Ram is 165 cms and of Shyam is 155 cms.
- (iii) Ram is taller than Shyam.
- (iv) Ram is taller than Shyam by 10 cms.
- (v) The height of Ram is 165 cms and weight of Shyam is 55 kgs.

Solution: Each of the above statement should be examined with reference to the following conditions:

- (a) Whether information is presented as aggregate of numerical figures
- (b) Whether numerical figures are homogeneous or comparable
- (c) Whether numerical figures are affected by a multiplicity of factors

On examination of the given information in the light of these conditions we find that only the information given by statement (ii) can be regarded as statistics.

It should be noted that condition (c) will be satisfied, almost invariably. In order to illustrate the circumstances in which this condition is not satisfied, we assume that a relation between quantity demanded and price of a commodity is given by the mathematical equation $q = 100 - 10p$ and the quantity demanded at various prices, using this equation, is shown in the following table,

p	1	2	3	4	5	6	7	8	9	10
q	90	80	70	60	50	40	30	20	10	0

The above information cannot be regarded as statistics because here quantity demanded is affected by only one factor, i.e., price and not by a multiplicity of factors. Contrary to this, the figures of quantity demanded obtained from a market at these very prices are to be regarded as statistics.

1.7.4 Statistics as a Science

The use of the word 'STATISTICS' in singular form refers to a science which provides methods of collection, analysis and interpretation of statistical data. Thus, statistics as a science is defined on the basis of its functions and different scholars have defined it in a different way. In order to know about various aspects of statistics, we now state some of these definitions.

- 1. "Statistics is the science of counting." - **A.L. Bowley**
- 2. "Statistics may rightly be called the science of averages." - **A.L. Bowley**
- 3. "Statistics is the science of measurement of social organism regarded as a whole in all its manifestations." - **A.L. Bowley**
- 4. "Statistics is the science of estimates and probabilities." - **Boddington**

All of the above definitions are incomplete in one sense or the other because each consider only one aspect of statistics. According to the first definition, statistics is the science of counting. However, we know that if the population or group under investigation is large, we do not count but obtain estimates.

The second definition viz. statistics is the science of averages, covers only one aspect, i.e., measures of average but, besides this, there are other measures used to describe a given set of data.

The third definition limits the scope of statistics to social sciences only. Bowley himself realised this limitation and admitted that scope of statistics is not confined to this area only.

The fourth definition considers yet another aspect of statistics. Although, use of estimates and probabilities have become very popular in modern statistics but there are other techniques, as well, which are also very important.

The following definitions covers some more but not all aspects of statistics.

- 5. "The science of statistics is the method of judging collective, natural or social phenomena from the results obtained by the analysis or enumeration or collection of estimates." - **W.I. King**
- 6. "Statistics or statistical method may be defined as collection, presentation, analysis and interpretation of numerical data." - **Croxton and Cowden**

This is a simple and comprehensive definition of statistics which implies that statistics is a scientific method.

7. "Statistics is a science which deals with collection, classification and tabulation of numerical facts as the basis for the explanation, description and comparison of phenomena." - **Lovitt**
8. "Statistics is the science which deals with the methods of collecting, classifying, presenting, comparing and interpreting numerical data collected to throw some light on any sphere of enquiry." - **Seligman**

The definitions given by Lovitt and Seligman are similar to the definition of Croxton and Cowden except that they regard statistics as a science while Croxton and Cowden has termed it as a scientific method.

With the development of the subject of statistics, the definitions of statistics given above have also become outdated. In the last few decades the discipline of drawing conclusions and making decisions under uncertainty has grown which is proving to be very helpful to decision-makers, particularly in the field of business. Although, various definitions have been given which include this aspect of statistics also, we shall now give a definition of statistics, given by Spiegel, to reflect this new dimension of statistics.

9. "Statistics is concerned with scientific method for collecting, organising, summarising, presenting and analysing data as well as drawing valid conclusions and making reasonable decisions on the basis of such analysis."

On the basis of the above definitions we can say that statistics, in singular sense, is a science which consists of various statistical methods that can be used for collection, classification, presentation and analysis of data relating to social, political, natural, economical, business or any other phenomena. The results of the analysis can be used further to draw valid conclusions and to make reasonable decisions in the face of uncertainty.

1.7.5 Statistics as a Science different from Natural Sciences

Science is a body of systematised knowledge developed by generalisations of relations based on the study of cause and effect. These generalised relations are also called the laws of science. For example, there are laws in physics, chemistry, statistics, mathematics, etc. It is obvious from this that statistics is also a science like any other natural science. The basic difference between statistics and other natural sciences lies in the difference in conditions under which its experiments are conducted. Where as the experiments in natural sciences are done in laboratory, under more or less controlled conditions, the experiments in statistics are conducted under uncontrolled conditions. Consider, for example, the collection of data regarding expenditure of households in a locality. There may be a large number of factors affecting expenditure and some of these factors might be different for different households.

Due to these reasons, statistics is often termed as a non-experimental science while natural sciences are termed as experimental sciences. We may note here that social sciences like economics, business, sociology, geography, political science, etc., belong to the category of non-experimental science and thus, the laws and methods of statistics can be used to understand and analyse the problems of these sciences also.

1.7.6 Statistics as a Scientific Method

We have seen above that, statistics as a non-experimental science can be used to study and analyse various problems of social sciences. It may, however, be pointed out that there may be situations even in natural sciences, where conducting of an experiment under hundred per cent controlled conditions is rather impossible. Statistics, under such conditions, finds its use in natural sciences, like physics, chemistry, etc.

In view of the uses of statistics in almost all the disciplines of natural as well as social sciences, it will be more appropriate to regard it as a scientific method rather than a science. Statistics as a scientific method can be divided into the following two categories: (a) Theoretical Statistics and (b) Applied Statistics

- (a) **Theoretical Statistics:** Theoretical statistics can be further sub-divided into the following three categories:
- (i) *Descriptive Statistics:* All those methods which are used for the collection, classification, tabulation, diagrammatic presentation of data and the methods of calculating average, dispersion, correlation and regression, index numbers, etc., are included in descriptive statistics.
 - (ii) *Inductive Statistics:* It includes all those methods which are used to make generalisations about a population on the basis of a sample. The techniques of forecasting are also included in inductive statistics.
 - (iii) *Inferential Statistics:* It includes all those methods which are used to test certain hypotheses regarding characteristics of a population.
- (b) **Applied Statistics:** It consists of the application of statistical methods to practical problems. Design of sample surveys, techniques of quality control, decision-making in business, etc., are included in applied statistics.

1.7.7 Statistics as a Science or an Art

We have seen above that statistics is a science. Now we shall examine whether it is an art or not. We know that science is a body of systematised knowledge. How this knowledge is to be used for solving a problem is work of an art. In addition to this, art also helps in achieving certain objectives and to identify merits and demerits of methods that could be used. Since statistics possesses all these characteristics, it may be reasonable to say that it is also an art.

Thus, we conclude that since statistical methods are systematic and have general applications, therefore, statistics is a science. Further since the successful application of these methods depends, to a considerable degree, on the skill and experience of a statistician, therefore, statistics is an art also.

1.8 LET US SUM UP

The changes in the structure of human organisation, perfection in various fields and introduction of decision had given birth to quantitative technique. The application of Quantitative Techniques methods helps in making decisions in such complicated situation. Evidently the primarily objective of Quantitative Techniques is to study the different components of an organisation by employing the methods of mathematical statistics in order to get the behaviour with greater degree of control on the system. In short, the objective of Quantitative Technique is to make available scientific basis to the decision-maker, for solving the problem involving the interaction of different components of the organisation by employing a team of scientists from distinguish disciplines, all working in concert for finding a solution which is in the best interest of organisation as a whole. The best solution thus obtained is known as optimal decision.

1.9 LESSON-END ACTIVITIES

1. Visit a nearby Nokia priority center as I hope it will reach your city. Analyse the functioning of the priority center and see which types of Quantitative Techniques could be more useful and applicable. For your convenience and even giving you the clue that if there are more customers in the priority center and service centers are not able to fulfil the requirements waiting line will be the best approach.

2. Why there is a need of statistics. Indicate one incidence of statistics application in your daily routine. How the statistics application had bring a paradigm shift.

1.10 KEYWORDS

Management science

Model

Analysis

Decision-making

Mathematical model

Algorithm

Problem

1.11 QUESTIONS FOR DISCUSSION

1. Write True or False against each statement:

- (a) Accurate data for input values are essential.
- (b) A factor is developed to suit the problem.
- (c) Key decision and objective of the problem must be identified.
- (d) The methodology helps us in studying the scientific method.
- (e) Model does not facilitates managers to take decisions.

2. Briefly comment on the following statements:

- (a) Scientific management has got wide scope.
- (b) Implementation involves translation/application of solutions.
- (c) A model is a mathematical representation of a problem situation.
- (d) It is necessary to clearly understand the problem situation.
- (e) Scientific management techniques are available to solve managerial problem.

3. Fill in the blanks:

- (a) Once the _____ in tested and validated, it is ready for implementation.
- (b) Quantitative factors are considered in _____
- (c) Managerial science had _____ the organisation.
- (d) Managerial criticism had become _____
- (e) Fredrich W. Taylor developed the _____ management principle.

4. Distinguish between the following:

- (a) Quantitative Techniques and Management.
- (b) Solving the model and validating the model translation.
- (c) Translation & Application.

1.12 TERMINAL QUESTIONS

1. How useful are the Quantitative Techniques in decision-making?
2. Give the areas of application where Quantitative Techniques can be applied.

3. Explain the methodology adopted in solving problems with the help of a flow chart diagram.
4. What is a model? Explain with a suitable example.
5. What is meant by validation of model?
6. Explain the advantages of modelling with the help of a short example.
7. Discuss the advantages and limitations of using results from a mathematical model to make decision as out operations.
8. What are different type of models used in management science.
9. What are some of the opportunities in management science?
10. What is implementation and why it is important?
11. What are some of sources of input data?
12. Briefly trace the history of management science.
13. What is the Quantitative Techniques process? Give several examples of this process.
14. Give a brief account of the origin and development of statistics.
15. Define statistics and discuss its relationship with natural and other sciences.
16. Distinguish between statistical methods and statistics. Discuss the scope and significance of the study of statistics.
17. Who gave the following definitions of statistics?
 - (i) “Statistics is the science of counting”.
(Bowley, Boddington, King, Saligman)
 - (ii) “Statistics is the science of estimates and probabilities”.
(Webster, Secrist, Boddington, Yule & Kendall)
 - (iii) “The science of statistics is the method of judging collective, natural or social phenomena from the results obtained by the analysis or enumeration or collection of estimates”.
(Achenwall, Marshall, W.I. King, Croxton & Cowden)
18. “Statistics are numerical statements of facts, but all facts stated numerically are not statistics”. Clarify this statement and point out briefly which numerical statements of facts are statistics.
19. Discuss briefly the utility of statistics in economic analysis and business.
20. Which of the following statements are true?
 - (a) Statistics is helpful in administration.
 - (b) Statistics is helpful in business.
 - (c) Statistics is helpful in economic analysis.
 - (d) Statistics is helpful in all of the above.
21. “Statistics are the straws out of which I like other economists have to make bricks”. Discuss.
22. “Science without statistics bear no fruit, statistics without science have no roots”. Explain the above statement.
23. “It is usually said that statistics is science and art both”. Do you agree with this statement? Discuss the scope of statistics.
24. Define ‘Statistics’ and explain briefly the divisions of the science of statistics.
25. “Statistics is not a science, it is a scientific method”. Discuss it critically and explain the scope of statistics.

26. Explain clearly the three meanings of the word 'Statistics' contained in the following statement :

"You compute statistics from statistics by statistics".

[**Hint** : Mean, standard deviation, etc., computed from a sample are also known as statistics.]

27. "Economics and statistics are twin sisters". Discuss.
28. Discuss the nature and scope of statistics. What are the fields of investigation and research where statistical methods and techniques can be usefully employed?
29. Explain the following statements :
- "Statistics is the science of counting".
 - "Statistics is the science of estimates and probabilities".
 - "Statistics is the science of averages".
30. Explain by giving reasons whether the following are data or not:
- Arun is more intelligent than Avinash.
 - Arun got 75% marks in B.Sc. and Avinash got 70% marks in B.Com.
 - Arun was born on August 25, 1974.
 - The consumption function of a community is $C = 1,000 + 0.8Y$, therefore, the levels of consumption for different levels of income are :

Y	0	1000	2000	4000	6000	8000
C	1000	1800	2600	4200	5800	7400

31. "Statistics are aggregates of facts, affected to a marked extent by a multiplicity of causes".

Discuss the above statement and explain the main characteristics of statistics.

32. "Statistics are not merely heap of numbers". Explain.
33. Elucidate the following statement :

"Not a datum, but data are the subject-matter of statistics".

1.13 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

- (a) True (b) False (c) True (d) True (e) False
- (a) model (b) decision-making (c) facilitate (d) complex (e) scientific

1.14 SUGGESTED READINGS

Bierman & Hausman, *Quantitative Analysis for Business Decision*.

Billy E Gillert, "*Introduction to OR*"

Franklyn A Lindsay, "*New Technique for Management Decision-making*".

Herbert G. Hicks, "*New Management of Organisation*".

Joseph L. Massie, "*Essentials of Management*."

R. L. Acnoff & M. W. Sasieni, "*Fundamentals of OR*".

Norbert Lloyd, Enrich, "*Management OR*".

LESSON

2

MEASURES OF CENTRAL TENDENCY

CONTENTS

- 2.0 Aims and Objectives
 - 2.1 Introduction
 - 2.2 Definition of Average
 - 2.3 Functions and Characteristics of an Average
 - 2.4 Various Measures of Average
 - 2.5 Arithmetic Mean
 - 2.6 Median
 - 2.7 Other Partition or Positional Measures
 - 2.8 Mode
 - 2.9 Relation between Mean, Median and Mode
 - 2.10 Geometric Mean
 - 2.11 Harmonic Mean
 - 2.12 Let us Sum Up
 - 2.13 Lesson-end Activity
 - 2.14 Keywords
 - 2.15 Questions for Discussion
 - 2.16 Terminal Questions
 - 2.17 Model Answers to Questions for Discussion
 - 2.18 Suggested Readings
-

2.0 AIMS AND OBJECTIVES

In this lesson we would be able to measure the various measures of Central Tendency like Average, Arithmetic mean, Median, Mode and the relationship between various measures of tendencies. We would also learn the Geometric and Harmonic Mean.

2.1 INTRODUCTION

Summarisation of the data is a necessary function of any statistical analysis. As a first step in this direction, the huge mass of unwieldy data are summarised in the form of tables and frequency distributions. In order to bring the characteristics of the data into sharp focus, these tables and frequency distributions need to be summarised further. A measure of central tendency or an average is very essential and an important summary measure in any statistical analysis. An average is a single value which can be taken as representative of the whole distribution.

2.2 DEFINITION OF AVERAGE

The average of a distribution has been defined in various ways. Some of the important definitions are :

- (i) "An average is an attempt to find one single figure to describe the whole of figures".
- Clark and Sekkade
- (ii) "Average is a value which is typical or representative of a set of data".
- Murray R. Spiegel
- (iii) "An average is a single value within the range of the data that is used to represent all the values in the series. Since an average is somewhere within the range of data it is sometimes called a measure of central value".
- Croxton and Cowden
- (iv) "A measure of central tendency is a typical value around which other figures congregate".
- Sipson and Kafka

2.3 FUNCTIONS AND CHARACTERISTICS OF AN AVERAGE

1. **To present huge mass of data in a summarised form:** It is very difficult for human mind to grasp a large body of numerical figures. A measure of average is used to summarise such data into a single figure which makes it easier to understand and remember.
2. **To facilitate comparison:** Different sets of data can be compared by comparing their averages. For example, the level of wages of workers in two factories can be compared by mean (or average) wages of workers in each of them.
3. **To help in decision-making:** Most of the decisions to be taken in research, planning, etc., are based on the average value of certain variables. For example, if the average monthly sales of a company are falling, the sales manager may have to take certain decisions to improve it.

Characteristics of a Good Average

A good measure of average must possess the following characteristics :

1. It should be rigidly defined, preferably by an algebraic formula, so that different persons obtain the same value for a given set of data.
2. It should be easy to compute.
3. It should be easy to understand.
4. It should be based on all the observations.
5. It should be capable of further algebraic treatment.
6. It should not be unduly affected by extreme observations.
7. It should not be much affected by the fluctuations of sampling.

2.4 VARIOUS MEASURES OF AVERAGE

Various measures of average can be classified into the following three categories:

- (a) Mathematical Averages :
 - (i) Arithmetic Mean or Mean
 - (ii) Geometric Mean
 - (iii) Harmonic Mean
 - (iv) Quadratic Mean
- (b) Positional Averages:
 - (i) Median
 - (ii) Mode

- (c) Commercial Average
 - (i) Moving Average
 - (ii) Progressive Average
 - (iii) Composite Average

The above measures of central tendency will be discussed in the order of their popularity. Out of these, the Arithmetic Mean, Median and Mode, being most popular, are discussed in that order.

2.5 ARITHMETIC MEAN

Before the discussion of arithmetic mean, we shall introduce certain notations. It will be assumed that there are n observations whose values are denoted by X_1, X_2, \dots, X_n respectively. The sum of these observations $X_1 + X_2 + \dots + X_n$ will be denoted in

abbreviated form as $\sum_{i=1}^n X_i$, where S (called sigma) denotes summation sign.

The subscript of X , i.e., ' i ' is a positive integer, which indicates the serial number of the observation. Since there are n observations, variation in i will be from 1 to n . This is indicated by writing it below and above S , as written earlier. When there is no ambiguity in range of summation, this indication can be skipped and we may simply write $X_1 + X_2 + \dots + X_n = \Sigma X_i$.

Arithmetic Mean is defined as the sum of observations divided by the number of observations. It can be computed in two ways : (i) Simple arithmetic mean and (ii) weighted arithmetic mean. In case of simple arithmetic mean, equal importance is given to all the observations while in weighted arithmetic mean, the importance given to various observations is not same.

Calculation of Simple Arithmetic Mean

(a) When Individual Observations are given.

Let there be n observations X_1, X_2, \dots, X_n . Their arithmetic mean can be calculated either by direct method or by short cut method. The arithmetic mean of these observations will be denoted by \bar{X}

Direct Method: Under this method, \bar{X} is obtained by dividing sum of observations by number of observations, i.e.,

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

Short-cut Method: This method is used when the magnitude of individual observations is large. The use of short-cut method is helpful in the simplification of calculation work.

Let A be any assumed mean. We subtract A from every observation. The difference between an observation and A , i.e., $X_i - A$ is called the deviation of i th observation from A and is denoted by d_i . Thus, we can write ; $d_1 = X_1 - A$, $d_2 = X_2 - A$, ..., $d_n = X_n - A$. On adding these deviations and dividing by n we get

$$\frac{\sum d_i}{n} = \frac{\sum (X_i - A)}{n} = \frac{\sum X_i - nA}{n} = \frac{\sum X_i}{n} - A$$

$$\text{or } \bar{d} = \bar{X} - A \quad (\text{Where } \bar{d} = \frac{\sum d_i}{n})$$

$$\text{On rearranging, we get } \bar{X} = A + \bar{d} = A + \frac{\sum d_i}{n}$$

This result can be used for the calculation of \bar{X} .

Remarks: Theoretically we can select any value as assumed mean. However, for the purpose of simplification of calculation work, the selected value should be as nearer to the value of \bar{X} as possible.

Example 1: The following figures relate to monthly output of cloth of a factory in a given year:

Months	: Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Output (in '000 metres)	: 80	88	92	84	96	92	96	100	92	94	98	86

Calculate the average monthly output.

Solution:

(i) **Using Direct Method**

$$\bar{X} = \frac{80 + 88 + 92 + 84 + 96 + 92 + 96 + 100 + 92 + 94 + 98 + 86}{12} = 91.5 \text{ ('000 mtrs)}$$

(ii) **Using Short Cut Method**

Let $A = 90$.

X_i	80	88	92	84	96	92	96	100	92	94	98	86	Total
$d_i = X_i - A$	-10	-2	2	-6	6	2	6	10	2	4	8	-4	$\sum d_i = 18$

$$\therefore \bar{X} = 90 + \frac{18}{12} = 90 + 1.5 = 91.5 \text{ thousand mtrs}$$

(b) **When data are in the form of an ungrouped frequency distribution**

Let there be n values X_1, X_2, \dots, X_n out of which X_1 has occurred f_1 times, X_2 has occurred f_2 times, X_n has occurred f_n times. Let N be the total frequency,

i.e., $N = \sum_{i=1}^n f_i$. Alternatively, this can be written as follows :

Values	X_1	X_2	-	-	X_n	Total Frequency
Frequency	f_1	f_2	-	-	f_n	N

Direct Method : The arithmetic mean of these observations using direct method is given by

$$x = \frac{\underbrace{X_1 + X_1 + \dots + X_1}_{f_1 \text{ times}} + \underbrace{X_2 + \dots + \dots + X_2 + \dots + \dots + X_n + \dots + X_n}_{f_2 \text{ times}} + \dots + \underbrace{X_n + \dots + X_n}_{f_n \text{ times}}}{f_1 + f_2 + \dots + f_n}$$

Since $X_1 + X_1 + \dots + X_1$ added f_1 times can also be written $f_1 X_1$. Similarly, by writing other observation in same manner, we have

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + \dots + f_n X_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum_{i=1}^n f_i X_i}{\sum_{i=1}^n f_i} = \frac{\sum_{i=1}^n f_i X_i}{N} \quad \dots (3)$$

Short-Cut Method: As before, we take the deviations of observations from an arbitrary value A . The deviation of i^{th} observation from A is $d_i = X_i - A$.

Multiplying both sides by f_i we have $f_i d_i = f_i (X_i - A)$

Taking sum over all the observations

$$\sum f_i d_i = \sum f_i (X_i - A) = \sum f_i X_i - A \sum f_i = \sum f_i X_i - A \cdot N$$

Dividing both sides by N we have

$$\frac{\sum f_i d_i}{N} = \frac{\sum f_i X_i}{N} - A = \bar{X} - A \text{ or } \bar{X} = A + \frac{\sum f_i d_i}{N} = A + \bar{d}.$$

Example 2: The following is the frequency distribution of age of 670 students of a school. Compute the arithmetic mean of the data.

X (in years)	5	6	7	8	9	10	11	12	13	14
Frequency	25	45	90	165	112	96	81	26	18	12

Solution:

Direct method: The computations are shown in the following table :

X	5	6	7	8	9	10	11	12	13	14	Total
f	25	45	90	165	112	96	81	26	18	12	$\sum f = 670$
fX	125	270	630	1320	1008	960	891	312	234	168	$\sum fX = 5918$

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{5918}{670} = 8.83 \text{ years.}$$

Short-Cut Method: The method of computations are shown in the following table :

X	5	6	7	8	9	10	11	12	13	14	Total
f	25	45	90	165	112	96	81	26	18	12	670
$d = X - 8$	-3	-2	-1	0	1	2	3	4	5	6	
fd	-75	-90	-90	0	112	192	243	104	90	72	558

$$\therefore \bar{X} = A + \frac{\sum fd}{N} = 8 + \frac{558}{670} = 8 + 0.83 = 8.83 \text{ years.}$$

(c) When data are in the form of a grouped frequency distribution

In a grouped frequency distribution, there are classes along with their respective frequencies. Let l_i be the lower limit and u_i be the upper limit of i th class. Further, let the number of classes be n , so that $i = 1, 2, \dots, n$. Also let f_i be the frequency of i th class. This distribution can be written in tabular form, as shown.

Note: Here u_i may or may not be equal to l_{i+1} , i.e., the upper limit of a class may or may not be equal to the lower limit of its following class.

It may be recalled here that, in a grouped frequency distribution, we only know the number of observations in a particular class interval and not their individual magnitudes. Therefore, to calculate mean, we have to make a fundamental assumption that the observations in a class are uniformly distributed. Under this assumption, the mid-value of a class will be equal to the mean of observations in that class and hence can be taken as their representative. Therefore, if X_i is the mid-value of i th class with frequency f_i , the above assumption implies that there are f_i observations each with magnitude X_i ($i = 1$ to n). Thus, the arithmetic mean of a grouped frequency distribution can also be calculated by the use of the formula, given in § 9.5.1(b).

Class Intervals	Frequency (f)
$l_1 - u_1$	f_1
$l_2 - u_2$	f_2
\vdots	\vdots
$l_n - u_n$	f_n
Total Frequency	$= \sum f_i = N$

Remarks: The accuracy of arithmetic mean calculated for a grouped frequency distribution depends upon the validity of the fundamental assumption. This assumption is rarely met in practice. Therefore, we can only get an approximate value of the arithmetic mean of a grouped frequency distribution.

Example 3: Calculate arithmetic mean of the following distribution :

<i>Class Intervals</i>	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
<i>Frequency</i>	3	8	12	15	18	16	11	5

Solution: Here only short-cut method will be used to calculate arithmetic mean but it can also be calculated by the use of direct-method.

<i>Class Intervals</i>	<i>Mid Values (X)</i>	<i>Frequency (f)</i>	$d = X - 35$	fd
0-10	5	3	-30	-90
10-20	15	8	-20	-160
20-30	25	12	-10	-120
30-40	35	15	0	0
40-50	45	18	10	180
50-60	55	16	20	320
60-70	65	11	30	330
70-80	75	5	40	200
<i>Total</i>		88		660

$$\therefore \bar{X} = A + \frac{\sum fd}{N} = 35 + \frac{660}{88} = 42.5$$

Example 4: The following table gives the distribution of weekly wages of workers in a factory. Calculate the arithmetic mean of the distribution.

<i>Weekly Wages</i>	240 - 269	270 - 299	300 - 329	330 - 359	360 - 389	390 - 419	420 - 449
<i>No. of Workers</i>	7	19	27	15	12	12	8

Solution: It may be noted here that the given class intervals are inclusive. However, for the computation of mean, they need not be converted into exclusive class intervals.

<i>Class Intervals</i>	<i>Mid Values (X)</i>	<i>Frequency</i>	$d = X - 344.5$	fd
240-269	254.5	7	-90	-630
270-299	284.5	19	-60	-1140
300-329	314.5	27	-30	-810
330-359	344.5	15	0	0
360-389	374.5	12	30	360
390-419	404.5	12	60	720
420-449	434.5	8	90	720
<i>Total</i>		100		-780

$$\therefore \bar{X} = A + \frac{\sum fd}{N} = 344.5 - \frac{780}{100} = 336.7$$

Step deviation method or coding method

In a grouped frequency distribution, if all the classes are of equal width, say 'h', the successive mid-values of various classes will differ from each other by this width. This fact can be utilised for reducing the work of computations.

Let us define $u_i = \frac{X_i - A}{h}$. Multiplying both sides by f_i and taking sum over all the

observations we have, $\sum_{i=1}^n f_i u_i = \frac{1}{h} \sum_{i=1}^n f_i (X_i - A)$

$$\text{or } h \sum_{i=1}^n f_i u_i = \sum_{i=1}^n f_i X_i - A \sum_{i=1}^n f_i = \sum_{i=1}^n f_i X_i - A.N$$

Dividing both sides by N, we have

$$h \cdot \frac{\sum_{i=1}^n f_i u_i}{N} = \frac{\sum_{i=1}^n f_i X_i}{N} - A = \bar{X} - A$$

$$\therefore \bar{X} = A + h \cdot \frac{\sum_{i=1}^n f_i u_i}{N} \quad \dots (5)$$

Using this relation we can simplify the computations of Example 4, as shown below.

$u = \frac{X - 344.5}{30}$	-3	-2	-1	0	1	2	3	Total
f	7	19	27	15	12	12	8	100
fu	-21	-38	-27	0	12	24	24	-26

Using formula (5), we have

$$\bar{X} = 344.5 - \frac{30 \times 26}{100} = 336.7$$

Example 5: Following table gives the distribution of companies according to size of capital. Find the mean size of the capital of a company.

Capital (Lacs Rs)	<5	<10	<15	<20	<25	<30
No. of Companies	20	27	29	38	48	53

Solution: This is a 'less than' cumulative frequency distribution. This will first be converted into class intervals.

Class Intervals	Frequency (f)	Mid-values (X)	$u = \frac{X - 12.5}{5}$	fu
0-5	20	2.5	-2	-40
5-10	7	7.5	-1	-7
10-15	2	12.5	0	0
15-20	9	17.5	1	9
20-25	10	22.5	2	20
25-30	5	27.5	3	15
Total	53			-3

$$\therefore \bar{X} = 12.5 - \frac{5 \times 3}{53} = \text{Rs } 12.22 \text{ Lacs}$$

Example 6: A charitable organisation decided to give old age pension to people over sixty years of age. The scale of pension were fixed as follows :

Age Group	: 60-65	65-70	70-75	75-80	80-85	85-90	90-95
Pension/Month (Rs)	: 100	120	140	160	180	200	220

If the total pension paid per month in various age groups are :

Age Group	: 60-65	65-70	70-75	75-80	80-85	85-90	90-95
Total Pension / Month	: 700	600	840	800	720	600	440

Calculate the average amount of pension paid per month per head and the average age of the group of old persons.

Solution: The computations of pension per head and the average age are shown in the following table.

Age Group	Rate of Pension per month (Y) (in Rs)	Total Pension paid per month (T) (in Rs)	No. of Persons f=T/Y	Mid-values of Class Intervals (X)	$u = \frac{X - 77.5}{5}$	fu
60-65	100	700	7	62.5	-3	-21
65-70	120	600	5	67.5	-2	-10
70-75	140	840	6	72.5	-1	-6
75-80	160	800	5	77.5	0	0
80-85	180	720	4	82.5	1	4
85-90	200	600	3	87.5	2	6
90-95	220	440	2	92.5	3	6
Total		4700	32			-21

$$\text{Average age } \bar{X} = 77.5 + \frac{5 \times (-21)}{32} = 77.5 - 3.28 = 74.22 \text{ Years}$$

$$\text{The average pension per head} = \frac{\text{Total pension paid}}{\text{No. of persons}} = \frac{4700}{32} = \text{Rs } 146.88$$

Charlier's Check of Accuracy

When the arithmetic mean of a frequency distribution is calculated by short-cut or step-deviation method, the accuracy of the calculations can be checked by using the following formulae, given by Charlier.

For short-cut method

$$\begin{aligned} \sum f_i(d_i + 1) &= \sum f_i d_i + \sum f_i \\ \text{or} \quad \sum f_i d_i &= \sum f_i(d_i + 1) - \sum f_i = \sum f_i(d_i + 1) - N \end{aligned}$$

Similarly, for step-deviation method

$$\begin{aligned} \sum f_i(u_i + 1) &= \sum f_i u_i + \sum f_i \\ \text{or} \quad \sum f_i u_i &= \sum f_i(u_i + 1) - \sum f_i = \sum f_i(u_i + 1) - N \end{aligned}$$

Checking the accuracy of calculations of Example 5, we have

$$\sum f(u+1) = 20 \times (-1) + (7 \times 0) + (2 \times 1) + (9 \times 2) + (10 \times 3) + (5 \times 4) = 50$$

Since $\sum f(u+1) - N = 50 - 53 = -3 = \sum fu$, the calculations are correct.

Weighted Arithmetic Mean

In the computation of simple arithmetic mean, equal importance is given to all the items. But this may not be so in all situations. If all the items are not of equal importance, then simple arithmetic mean will not be a good representative of the given data. Hence, weighing of different items becomes necessary. The weights are assigned to different items depending upon their importance, i.e., more important items are assigned more weight. For example, to calculate mean wage of the workers of a factory, it would be wrong to compute simple arithmetic mean if there are a few workers (say managers) with very high wages while majority of the workers are at low level of wages. The simple arithmetic mean, in such a situation, will give a higher value that cannot be regarded as representative wage for the group. In order that the mean wage gives a realistic picture of the distribution, the wages of managers should be given less importance in its computation. The mean calculated in this manner is called weighted arithmetic mean. The computation of weighted arithmetic is useful in many situations where different items are of unequal importance, e.g., the construction index numbers, computation of standardised death and birth rates, etc.

Formulae for Weighted Arithmetic Mean

Let X_1, X_2, \dots, X_n be n values with their respective weights w_1, w_2, \dots, w_n . Their weighted arithmetic mean denoted as \bar{X}_w is given by,

$$(i) \quad \bar{X}_w = \frac{\sum w_i X_i}{\sum w_i} \quad (\text{Using direct method}),$$

$$(ii) \quad \bar{X}_w = A + \frac{\sum w_i d_i}{\sum w_i} \quad (\text{where } d_i = X_i - A) \quad (\text{Using short-cut method}),$$

$$(iii) \quad \bar{X}_w = A + \frac{\sum w_i u_i}{\sum w_i} \times h \quad (\text{where } u_i = \frac{X_i - A}{h}) \quad (\text{Using step-deviation method})$$

Example 7: Ram purchased equity shares of a company in 4 successive months, as given below. Find the average price per share.

Month	No. of Shares	Price per share (in Rs.)
Dec - 91	100	200
Jan - 92	150	250
Feb - 92	200	280
Mar - 92	125	300

Solution: The average price is given by the weighted average of prices, taking the number of shares purchased as weights.

Month	Price of share (X) (in Rs)	No. of shares (w)	$d = X - 150$	dw
Dec - 91	100	200	-50	-10000
Jan - 92	150	250	0	0
Feb - 92	200	280	50	14000
Mar - 92	125	300	-25	-7500
Total		1030		-3500

$$\bar{X}_w = 150 - \frac{3,500}{1,030} = \text{Rs } 146.60$$

Example 8: From the following results of two colleges A and B, find out which of the two is better :

Examination	College A		College B	
	Appeared	Passed	Appeared	Passed
M.Sc.	60	40	200	160
M.A.	100	60	240	200
B.Sc.	200	150	200	140
B.A.	120	75	160	100

Solution: Performance of the two colleges can be compared by taking weighted arithmetic mean of the pass percentage in various classes. The calculation of weighted arithmetic mean is shown in the following table.

Class	College A				College B			
	Appeared w_A	Passed	Pass Percentage X_A	$w_A X_A$	Appeared w_B	Passed	Pass Percentage X_B	$w_B X_B$
M.Sc.	60	40	66.67	4000.2	200	160	80.00	16000.0
M.A.	100	60	60.00	6000.0	240	200	83.33	19999.2
B.Sc.	200	150	75.00	15000.0	200	140	70.00	14000.0
B.A.	120	75	62.50	7500.0	160	100	62.50	10000.0
Total	480	325		32500.2	800	600		59999.2

$$\bar{X}_w \text{ for College A} = \frac{\sum w_A X_A}{\sum w_A} = \frac{32500.2}{480} = 67.71\%$$

$$\bar{X}_w \text{ for College B} = \frac{\sum w_B X_B}{\sum w_B} = \frac{59999.2}{800} = 75\%$$

Since the weighted average of pass percentage is higher for college B, hence college B is better.

Remarks: If \bar{X} denotes simple mean and \bar{X}_w denotes the weighted mean of the same data, then

- (i) $\bar{X} = \bar{X}_w$, when equal weights are assigned to all the items.
- (ii) $\bar{X} > \bar{X}_w$, when items of small magnitude are assigned greater weights and items of large magnitude are assigned lesser weights.
- (iii) $\bar{X} < \bar{X}_w$, when items of small magnitude are assigned lesser weights and items of large magnitude are assigned greater weights.

Properties of Arithmetic Mean

Arithmetic mean of a given data possess the following properties :

1. The sum of deviations of the observations from their arithmetic mean is always zero.

According to this property, the arithmetic mean serves as a point of balance or a centre of gravity of the distribution; since sum of positive deviations (i.e., deviations of observations which are greater than \bar{X}) is equal to the sum of negative deviations (i.e., deviations of observations which are less than \bar{X}).

Proof: Let X_1, X_2, \dots, X_n be n observations with respective frequencies f_1, f_2, \dots, f_n . Let $Sf_i = N$, be the total frequency. Thus, $\bar{X} = \frac{\sum f_i X_i}{N}$

Let $d_i = X_i - \bar{X}$, where $i = 1, 2, \dots, n$. Multiplying both sides by f_i and taking sum over all the observations, we have

$$\begin{aligned}\sum f_i d_i &= \sum f_i (X_i - \bar{X}) = \sum f_i X_i - \bar{X} \sum f_i \\ &= \sum f_i X_i - \bar{X} \cdot N = 0 \quad (\text{Since } \sum f_i X_i = N\bar{X}). \text{ Hence Proved.}\end{aligned}$$

2. The sum of squares of deviations of observations is minimum when taken from their arithmetic mean. Because of this, the mean is sometimes termed as 'least square' measure of central tendency.

Proof: The sum of squares of deviations of observations from arithmetic mean

$$= \sum f_i (X_i - \bar{X})^2$$

Similarly, we can define sum of squares of deviations of observations from any arbitrary value A as $S = \sum f_i (X_i - A)^2$ (1)

We have to show that S will be minimum when $A = \bar{X}$.

To prove this, we try to find that value of A for which S is minimum. The necessary and sufficient conditions for minimum of S are :

$$\frac{dS}{dA} = 0 \quad \text{and} \quad \frac{d^2 S}{dA^2} > 0$$

Differentiating (1) w.r.t. A we have

$$\frac{dS}{dA} = -2 \sum f_i(X_i - A) = 0, \text{ for minima.} \quad \dots (2)$$

On dividing both sides by - 2, we have

$$\sum f_i(X_i - A) = 0 \quad \text{or} \quad \sum f_i X_i - N A = 0$$

$$\text{or } \frac{\sum f_i X_i}{N} - A = 0 \quad (\text{on dividing both sides by } N)$$

$$\text{Thus, } \bar{X} - A = 0 \quad \text{or} \quad A = \bar{X}.$$

Further, to show that S is minimum, it will be shown that $\frac{d^2S}{dA^2} > 0$ at $A = \bar{X}$.

Differentiating (2) further w.r.t. A, we have

$$\frac{d^2S}{dA^2} = -2 \sum f_i(0-1) = 2 \sum f_i = 2N, \text{ which is always positive.}$$

Hence S is minimum when $A = \bar{X}$.

3. Arithmetic mean is capable of being treated algebraically.

This property of arithmetic mean highlights the relationship between \bar{X} , $\sum f_i X_i$ and N. According to this property, if any two of the three values are known, the third can be easily computed. This property is obvious and requires no proof.

4. If \bar{X}_1 and N_1 are the mean and number of observations of a series and \bar{X}_2 and N_2 are the corresponding magnitudes of another series, then the mean \bar{X} of the combined series of $N_1 + N_2$ observations is given by $\bar{X} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2}$.

Proof: To find mean of the combined series, we have to find sum of its observations.

Now, the sum of observations of the first series, i.e., $\sum f_i X_i = N_1 \bar{X}_1$ and the sum of observations of the second series, i.e., $\sum f_i X_i = N_2 \bar{X}_2$.

\therefore The sum of observations of the combined series, i.e., $N_1 \bar{X}_1 + N_2 \bar{X}_2$.

$$\text{Thus, the combined mean } \bar{X} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2}.$$

This result can be generalised: If there are k series each with mean \bar{X}_i and number of observations equal to N_i , where $i = 1, 2, \dots, k$, the mean of the combined series of $N_1 + N_2 + \dots + N_k$ observations is given by

$$\bar{X} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2 + \dots + N_k \bar{X}_k}{N_1 + N_2 + \dots + N_k} = \frac{\sum_{i=1}^k N_i \bar{X}_i}{\sum_{i=1}^k N_i}$$

5. If a constant B is added (subtracted) from every observation, the mean of these observations also gets added (subtracted) by it.

Proof : Let \bar{X} be the mean of the observations X_1, X_2, \dots, X_n with respective frequencies as f_1, f_2, \dots, f_n . When B is added to every observations, let $u_i = X_i + B$.

Multiply both sides by f_i and take sum over all the observations, we get

$$\sum f_i u_i = \sum f_i (X_i + B) = \sum f_i X_i + NB$$

Dividing both sides by N we get

$$\frac{\sum f_i u_i}{N} = \frac{\sum f_i X_i}{N} + B \text{ or } \bar{u} = \bar{X} + B.$$

i.e., The mean of $u_i = X_i + B$ is obtained by adding B to the mean of X_i values.

Similarly, it can be shown that if $v_i = X_i - B$, then $\bar{v} = \bar{X} - B$.

6. If every observation is multiplied (divided) by a constant b , the mean of these observations also gets multiplied (divided) by it.

Proof: Let us define $w_i = \beta X_i$. Multiplying both sides by f_i and taking sum over all the observations, we get $\sum f_i w_i = \beta \sum f_i X_i$.

$$\text{Dividing both sides by } N, \text{ we get } \frac{\sum f_i w_i}{N} = \beta \frac{\sum f_i X_i}{N} \text{ or } \bar{w} = \beta \bar{X}$$

Similarly, it can be shown that if $D_i = \frac{X_i}{\beta}$, then $\bar{D} = \frac{\bar{X}}{\beta}$

Using properties 5 and 6, we can derive the following results :

If $Y_i = a + bX_i$, then $\sum f_i Y_i = \sum f_i (a + bX_i)$ or $\sum f_i Y_i = a \sum f_i + b \sum f_i X_i$.

Dividing both sides by $N (= \sum f_i)$, we have

$$\frac{\sum f_i Y_i}{N} = a + b \frac{\sum f_i X_i}{N} \text{ or } \bar{Y} = a + b \bar{X}$$

This shows that relationship between the means of two variables is same as the relationship between the variables themselves.

7. If some observations of a series are replaced by some other observations, then the mean of original observations will change by the average change in magnitude of the changed observations.

Proof: Let mean of n observations be $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$. Further, Let X_1, X_2, X_3 are replaced by the respective observations Y_1, Y_2, Y_3 . Therefore, the change in magnitude of the changed observations $= (Y_1 + Y_2 + Y_3) - (X_1 + X_2 + X_3)$.

$$\text{Hence average change in magnitude} = \frac{(Y_1 + Y_2 + Y_3) - (X_1 + X_2 + X_3)}{n}.$$

Thus, new \bar{X} = old \bar{X} + average change in magnitude.

Example 9: There are 130 teachers and 100 non-teaching employees in a college. The respective distributions of their monthly salaries are given in the following table :

Teachers		Non-Teaching Employees	
Monthly Salary (in Rs)	Frequency	Monthly Salary (in Rs)	Frequency
4000 - 5000	10	1000 - 2000	21
5000 - 6000	16	2000 - 3000	45
6000 - 7000	22	3000 - 4000	28
7000 - 8000	67	4000 - 5000	06
8000 - 9000	15		
Total	130	Total	100

From the above data find :

- (i) Average monthly salary of a teacher.
- (ii) Average monthly salary of a non-teaching employee.
- (iii) Average monthly salary of a college employee (teaching and non-teaching).

Solution:

- (i) Average monthly salary of a teacher

Monthly Salary (in Rs)	Frequency <i>f</i>	Mid-Values (X)	$u = \frac{X - 6500}{1000}$	fu
4000 - 5000	10	4500	-2	-20
5000 - 6000	16	5500	-1	-16
6000 - 7000	22	6500	0	0
7000 - 8000	67	7500	1	67
8000 - 9000	15	8500	2	30
Total	130			61

$$\therefore \bar{X}_1 = 6500 + 469.23 = \text{Rs } 6969.23$$

- (ii) Average monthly salary of a non-teaching employee

Monthly Salary (in Rs)	Frequency <i>f</i>	Mid-Values (X)	$v = \frac{X - 2500}{1000}$	fv
1000 - 2000	21	1500	-1	-21
2000 - 3000	45	2500	0	0
3000 - 4000	28	3500	1	28
4000 - 5000	6	4500	2	12
Total	100			19

$$\therefore \bar{X}_2 = 2500 + 190 = \text{Rs } 2690$$

- (iii) Combined average monthly salary is

$$\bar{X} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2}$$

Here $N_1 = 130$, $\bar{X}_1 = 6969.23$, $N_2 = 100$ and $\bar{X}_2 = 2690$

$$\therefore \bar{X} = \frac{130 \times 6969.23 + 100 \times 2690}{230} = \text{Rs } 5108.70$$

Example 10: The average rainfall for a week, excluding Sunday, was 10 cms. Due to heavy rainfall on Sunday, the average for the week rose to 15 cms. How much rainfall was on Sunday?

Solution: A week can be treated as composed of two groups: First group consisting of 6 days excluding Sunday for which $N_1 = 6$ and $\bar{X}_1 = 10$; the second group consisting of only Sunday for which $N_2 = 1$. Also, mean of this group will be equal to the observation itself. Let this be X . We have to determine the value of X .

We are also given $N = 7$ and $\bar{X}_1 = 15$ (for the whole week).

$$\therefore 15 = \frac{6 \times 10 + 1 \times X}{7} \text{ or } 60 + X = 15 \times 7$$

$\Rightarrow X = 105 - 60 = 45$ cms. Thus, the rainfall on Sunday was 45 cms.

Example 11: The mean age of the combined group of men and women is 30.5 years. If the mean age of the sub-group of men is 35 years and that of the sub-group of women is 25 years, find out percentage of men and women in the group.

Solution: Let x be the percentage of men in the combined group. Therefore, percentage of women $= 100 - x$.

We are given that \bar{X}_1 (men) $= 35$ years and \bar{X}_2 (women) $= 25$ years

Also \bar{X} (combined) $= 30.5$

$$30.5 = \frac{35x + 25(100 - x)}{x + 100 - x} = \frac{35x + 2500 - 25x}{100} \text{ or } 3050 = 10x + 2500$$

$\Rightarrow x = \frac{550}{10} = 55\%$. Thus, there are 55% men and 45% women in the group.

To find missing frequency or a missing value

Example 12: The following is the distribution of weights (in lbs.) of 60 students of a class:

Weights	:	93 - 97	98 - 102	103 - 107	108 - 112	113 - 117
No. of Students	:	2	5	12	?	14
Weights	:	118 - 122	123 - 127	128 - 132	Total	
No. of Students	:	?	3	1	60	

If the mean weight of the students is 110.917, find the missing frequencies.

Solution: Let f_1 be the frequency of the class 108-112. Then, the frequency of the class 118-122 is given by $60 - (2 + 5 + 12 + 14 + 3 + 1 + f_1) = 23 - f_1$

Writing this information in tabular form we have :

Weights (in lbs.)	No. of Students (f)	Mid-Values (X)	$u = \frac{X - 110}{5}$	fu
93-97	2	95	-3	-6
98-102	5	100	-2	-10
103-107	12	105	-1	-12
108-112	f_1	110	0	0
113-117	14	115	1	14
118-122	$23 - f_1$	120	2	$46 - 2f_1$
123-127	3	125	3	9
128-132	1	130	4	4
<i>Total</i>	60			$45 - 2f_1$

Using the formula for A.M., we can write $110.917 = 110 + \frac{(45 - 2f_1)5}{60}$

or $11.004 = 45 - 2f_1$ or $2f_1 = 33.996 = 34$ (approximately)

Thus, $f_1 = 17$ is the frequency of the class 108 - 112 and $23 - 17 = 6$ is the frequency of the class 118 - 122.

Example 13: Find out the missing item (x) of the following frequency distribution whose arithmetic mean is 11.37.

$X :$	5	7	(x)	11	13	16	20
$f :$	2	4	29	54	11	8	4

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{(5 \times 2) + (7 \times 4) + 29x + (11 \times 54) + (13 \times 11) + (16 \times 8) + (20 \times 4)}{112}$$

$$11.37 = \frac{10 + 28 + 29x + 594 + 143 + 128 + 80}{112} \text{ or } 11.37 \times 112 = 983 + 29x$$

$$\therefore x = \frac{290.44}{29} = 10.015 = 10 \text{ (approximately)}$$

Example 14: The arithmetic mean of 50 items of a series was calculated by a student as 20. However, it was later discovered that an item 25 was misread as 35. Find the correct value of mean.

Solution: $N = 50$ and $\bar{X} = 20 \therefore \Sigma X_i = 50 \times 20 = 1000$

$$\text{Thus } \Sigma X_{i \text{ (corrected)}} = 1000 + 25 - 35 = 990 \text{ and } \bar{X}_{\text{(corrected)}} = \frac{990}{50} = 19.8$$

Alternatively, using property 7 :

$$\bar{X}_{\text{new}} = \bar{X}_{\text{old}} + \text{average change in magnitude} = 20 - \frac{10}{50} = 20 - 0.2 = 19.8$$

Example 15: The sales of a balloon seller on seven days of a week are as given below:

Days	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Sales (in Rs)	100	150	125	140	160	200	250

If the profit is 20% of sales, find his average profit per day.

Solution: Let P denote profit and S denote sales, $\therefore P = \frac{20}{100} \times S$

Using property 6, we can write $\bar{P} = \frac{20}{100} \times \bar{S}$ or $\bar{P} = \frac{1}{5} \times \bar{S}$

$$\text{Now } \bar{S} = \frac{100 + 150 + 125 + 140 + 160 + 200 + 250}{7} = 160.71$$

$$\therefore \bar{P} = \frac{160.71}{5} = \text{Rs } 32.14$$

Hence, the average profit of the balloon seller is Rs 32.14 per day.

Alternatively, we can find profit of each day and take mean of these values.

Days	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Profit (in Rs)	20	30	25	28	32	40	50

$$\bar{P} = \frac{20 + 30 + 25 + 28 + 32 + 40 + 50}{7} = \text{Rs } 32.14$$

Merits and Demerits of Arithmetic Mean

Merits

Out of all averages arithmetic mean is the most popular average in statistics because of its merits given below:

1. Arithmetic mean is rigidly defined by an algebraic formula.
2. Calculation of arithmetic mean requires simple knowledge of addition, multiplication and division of numbers and hence, is easy to calculate. It is also simple to understand the meaning of arithmetic mean, e.g., the value per item or per unit, etc.
3. Calculation of arithmetic mean is based on all the observations and hence, it can be regarded as representative of the given data.
4. It is capable of being treated mathematically and hence, is widely used in statistical analysis.
5. Arithmetic mean can be computed even if the detailed distribution is not known but sum of observations and number of observations are known.
6. It is least affected by the fluctuations of sampling.
7. It represents the centre of gravity of the distribution because it balances the magnitudes of observations which are greater and less than it.
8. It provides a good basis for the comparison of two or more distributions.

Demerits

Although, arithmetic mean satisfies most of the properties of an ideal average, it has certain drawbacks and should be used with care. Some demerits of arithmetic mean are:

1. It can neither be determined by inspection nor by graphical location.
2. Arithmetic mean cannot be computed for a qualitative data; like data on intelligence, honesty, smoking habit, etc.
3. It is too much affected by extreme observations and hence, it does not adequately represent data consisting of some extreme observations.
4. The value of mean obtained for a data may not be an observation of the data and as such it is called a fictitious average.
5. Arithmetic mean cannot be computed when class intervals have open ends. To compute mean, some assumption regarding the width of class intervals is to be made.
6. In the absence of a complete distribution of observations the arithmetic mean may lead to fallacious conclusions. For example, there may be two entirely different distributions with same value of arithmetic mean.
7. Simple arithmetic mean gives greater importance to larger values and lesser importance to smaller values.

Exercise with Hints

1. The frequency distribution of weights in grams of mangoes of a given variety is given below. Calculate the arithmetic mean.

<i>Weights (in gms)</i>	<i>No. of Mangoes</i>	<i>Weights (in gms)</i>	<i>No. of Mangoes</i>
410-419	14	450-459	45
420-429	20	460-469	18
430-439	42	470-479	7
440-449	52		

Hint : Take the mid-value of a class as the mean of its limits and find arithmetic mean by the step-deviation method.

2. The following table gives the monthly income (in rupees) of families in a certain locality. By stating the necessary assumptions, calculate arithmetic mean of the distribution.

<i>Income</i> :	1000	1000 - 2000	2000 - 3000	3000 - 4000	4000 - 5000	5000
<i>No. of Families</i> :	100	1200	1450	250	70	30

Hint : This distribution is with open end classes. To calculate mean, it is to be assumed that the width of first class is same as the width of second class. On this assumption the lower limit of the first class will be 0. Similarly, it is assumed that the width of last class is equal to the width of last but one class. Therefore, the upper limit of the last class can be taken as 6,000.

3. Compute arithmetic mean of the following distribution of marks in Economics of 50 students.

<i>Marks more than</i>	<i>No. of Students</i>	<i>Marks more than</i>	<i>No. of Students</i>
0	50	50	15
10	46	60	8
20	40	70	3
30	33	80	0
40	25		

Hint: First convert the distribution into class intervals and then calculate \bar{X} .

4. The monthly profits, in '000 rupees, of 100 shops are distributed as follows:

<i>Profit per Shop</i> :	0 - 100	0 - 200	0 - 300	0 - 400	0 - 500	0 - 600
<i>No. of Shops</i> :	12	30	57	77	94	100

Find average profit per shop.

Hint: This is a less than type cumulative frequency distribution.

5. Typist A can type a letter in five minutes, typist B in ten minutes and typist C in fifteen minutes. What is the average number of letters typed per hour per typist?

Hint: In one hour, A will type 12 letters, B will type 6 letters and C will type 4 letters.

6. A taxi ride in Delhi costs Rs 5 for the first kilometre and Rs 3 for every additional kilometre travelled. The cost of each kilometre is incurred at the beginning of the kilometre so that the rider pays for the whole kilometre. What is the average cost of travelling $2\frac{3}{4}$ kilometres?

Hint: Total cost of travelling $2\frac{3}{4}$ kilometres = Rs 5 + 3 + 3 = Rs 11.

7. A company gave bonus to its employees. The rates of bonus in various salary groups are :

<i>Monthly Salary (in Rs)</i> :	1000-2000	2000-3000	3000-4000	4000-5000
<i>Rate of Bonus (in Rs)</i> :	2000	2500	3000	3500

The actual salaries of staff members are as given below :

1120, 1200, 1500, 4500, 4250, 3900, 3700, 3950, 3750, 2900, 2500, 1650, 1350, 4800, 3300, 3500, 1100, 1800, 2450, 2700, 3550, 2400, 2900, 2600, 2750, 2900, 2100, 2600, 2350, 2450, 2500, 2700, 3200, 3800, 3100.

Determine (i) Total amount of bonus paid and (ii) Average bonus paid per employee.

Hint: Find the frequencies of the classes from the given information.

8. Calculate arithmetic mean from the following distribution of weights of 100 students of a college. It is given that there is no student having weight below 90 lbs. and the total weight of persons in the highest class interval is 350 lbs.

<i>Weights</i>	: < 100	< 110	< 120	< 130	< 140	< 150	< 160	< 170	170 ≥
<i>Frequency</i>	: 3	5	23	45	66	85	95	98	2

Hint: Rearrange this in the form of frequency distribution by taking class intervals as 90 - 100, 100 - 110, etc.

9. By arranging the following information in the form of a frequency distribution, find arithmetic mean.

"In a group of companies 15%, 25%, 40% and 75% of them get profits less than Rs 6 lakhs, 10 lakhs, 14 lakhs and 20 lakhs respectively and 10% get Rs 30 lakhs or more but less than 40 lakhs."

Hint: Take class intervals as 0 - 6, 6 - 10, 10 - 14, 14 - 20, etc.

10. Find class intervals if the arithmetic mean of the following distribution is 38.2 and the assumed mean is equal to 40.

<i>Step deviations</i>	: 3	2	1	0	1	2	3
<i>Frequency</i>	: 8	14	18	28	17	10	5

Hint: Use the formula $\bar{X} = A + \frac{\sum fu}{N} \times h$ to find the class width h.

11. From the following data, calculate the mean rate of dividend obtainable to an investor holding shares of various companies as shown :

<i>Percentage Dividend</i>	: 30 - 40	20 - 30	10 - 20	0 - 10
<i>No. of Companies</i>	: 4	25	15	6
<i>Average no. of shares of each company held by the investor</i>	: 250	150	200	300

Hint: The no. of shares of each type = no. of companies × average no. of shares.

12. The mean weight of 150 students in a certain class is 60 kgs. The mean weight of boys in the class is 70 kgs and that of girls is 55 kgs. Find the number of girls and boys in the class.

Hint: Take n_1 as the no. of boys and $150 - n_1$ as the no. of girls.

13. The mean wage of 100 labourers working in a factory, running two shifts of 60 and 40 workers respectively, is Rs 38. The mean wage of 60 labourers working in the morning shift is Rs 40. Find the mean wage of 40 labourers working in the evening shift.

Hint: See example 10.

14. The mean of 25 items was calculated by a student as 20. If an item 13 is replaced by 30, find the changed value of mean.

Hint: See example 14.

15. The average daily price of share of a company from Monday to Friday was Rs 130. If the highest and lowest price during the week were Rs 200 and Rs 100 respectively, find average daily price when the highest and lowest price are not included.

Hint: See example 10.

16. The mean salary paid to 1000 employees of an establishment was found to be Rs 180.40. Later on, after disbursement of the salary, it was discovered that the salaries of two employees were wrongly recorded as Rs 297 and Rs 165 instead of Rs 197 and Rs 185. Find the correct arithmetic mean.

Hint: See example 14.

17. Find the missing frequencies of the following frequency distribution :

<i>Class Intervals</i>	60 - 65	65 - 70	70 - 75	75 - 80	80 - 85	85 - 90	90 - 95	95 - 100
<i>Frequency</i>	5	10	26	35	?	20	15	?

Hint: See example 12.

18. Marks obtained by students who passed a given examination are given below :

<i>Marks obtained (in percent)</i>	40-50	50-60	60-70	70-80	80-90	90-100
<i>No. of Students</i>	10	12	20	9	5	4

If 100 students took the examination and their mean marks were 51, calculate the mean marks of students who failed.

Hint: See example 9.

19. A appeared in three tests of the value of 20, 50 and 30 marks respectively. He obtained 75% marks in the first and 60% marks in the second test. What should be his percentage of marks in the third test in order that his aggregate is 60%?

Hint: Let x be the percentage of marks in third test. Then the weighted average of 75, 60 and x should be 60, where weights are 20, 50 and 30 respectively.

20. Price of a banana is 80 paise and the price of an orange is Rs 1.20. If a person purchases two dozens of bananas and one dozen of oranges, show by stating reasons that the average price per piece of fruit is 93 paise and not one rupee.

Hint: Correct average is weighted arithmetic average.

21. The average marks of 39 students of a class is 50. The marks obtained by 40th student are 39 more than the average marks of all the 40 students. Find mean marks of all the 40 students.

Hint: $\bar{X} + 39 + 39 \times 50 = 40\bar{X}$.

22. The means calculated for frequency distributions I and II were 36 and 32 respectively. Find the missing frequencies of the two distributions.

<i>Class Intervals</i>	<i>Frequency of Distribution I</i>	<i>Frequency of Distribution II</i>
5 - 15	4	10
15 - 25	10	14
25 - 35	14	3y
35 - 45	16	13
45 - 55	2x	10
55 - 65	y	x

Hint: $36 = \frac{40 + 200 + 420 + 640 + 100x + 60y}{44 + 2x + y}$ and

$$32 = \frac{100 + 280 + 90y + 520 + 500 + 60x}{47 + 3y + x}$$

Solve these equations simultaneously for the values of x and y .

23. The following table gives the number of workers and total wages paid in three departments of a manufacturing unit :

<i>Department</i>	<i>No. of Workers</i>	<i>Total wages (in Rs)</i>
A	105	1,68,000
B	304	4,25,600
C	424	5,08,800

If a bonus of Rs 200 is given to each worker, what is the average percentage increase in wages of the workers of each department and of the total workers?

Hint: (i) Average wage in deptt. A = $\frac{1,68,000}{105} = 1,600$

$$\therefore \text{Percentage increase in wages} = \frac{200}{1,600} \times 100 = 12.5\%$$

Similarly, find for deptt. B and C.

$$(ii) \text{Average wage for total workers} = \frac{1,68,000 + 4,25,600 + 5,08,800}{105 + 304 + 424}$$

Then, find percentage increase as before.

24. The following table gives the distribution of the number of kilometres travelled per salesman, of a pharmaceutical company, per day and their rates of conveyance allowance:

No. of kilometre travelled per salesman	No. of salesman	Rate of conveyance allowance per kilometre (in Rs)
10 - 20	3	2.50
20 - 30	8	2.60
30 - 40	15	2.70
40 - 50	4	2.80

Calculate the average rate of conveyance allowance given to each salesman per kilometre by the company.

Hint: Obtain total number of kilometre travelled for each rate of conveyance allowance by multiplying mid-values of column 1 with column 2. Treat this as frequency 'f' and third column as 'X' and find \bar{X} .

25. The details of monthly income and expenditure of a group of five families are given in the following table:

Family	Income (in Rs)	Expenditure per member (in Rs)	No. of members in the family
A	1100	220	4
B	1200	190	5
C	1300	230	4
D	1400	260	3
E	1500	250	4

Find: (i) Average income per member for the entire group of families.
(ii) Average expenditure per family.
(iii) The difference between actual and average expenditure for each family.

Hint: (i) Average income per member = $\frac{\text{Total income of the group of families}}{\text{Total no. of members in the group}}$.

(ii) Average expenditure per family = $\frac{\text{Total expenditure of the group}}{\text{No. of families}}$.

26. The following table gives distribution of monthly incomes of 200 employees of a firm:

Income (in Rs '00)	: 10-15	15-20	20-25	25-30	30-35	35-40
No. of employees	: 30	50	55	32	20	13

Estimate:

- (i) Mean income of an employee per month.
(ii) Monthly contribution to welfare fund if every employee belonging to the top 80% of the earners is supposed to contribute 2% of his income to this fund.

Hint: The distribution of top 80% of the wage earners can be written as :

<i>Income (in Rs '00)</i>	: 16-20	20-25	25-30	30-35	35-40
<i>Frequency</i>	: 40	55	32	20	13

By taking mid-values of class intervals find S_{fx} , i.e., total salary and take 2% of this.

27. The number of patients visiting diabetic clinic and protein urea clinic in a hospital during April 1991, are given below :

<i>No. of Patients</i>	<i>No. of days of attending</i>	
	<i>Diabetic Clinic</i>	<i>Protein Urea Clinic</i>
0 - 10	2	4
10 - 20	8	6
20 - 30	7	5
30 - 40	7	8
40 - 50	4	3
50 - 60	2	4

Which of these two diseases has more incidence in April 1991? Justify your conclusion.

Hint: The more incidence of disease is given by higher average number of patients.

28. A company has three categories of workers A, B and C. During 1994, the number of workers in respective category were 40, 240 and 120 with monthly wages Rs 1,000, Rs 1,300 and Rs 1,500. During the following year, the monthly wages of all the workers were increased by 15% and their number, in each category, were 130, 150 and 20, respectively.
- (a) Compute the average monthly wages of workers for the two years.
 - (b) Compute the percentage change of average wage in 1995 as compared with 1994. Is it equal to 15%? Explain.

Hint: Since the weight of the largest wage is less in 1995, the increase in average wage will be less than 15%.

29. (a) The average cost of producing 10 units is Rs 6 and the average cost of producing 11 units is Rs 6.5. Find the marginal cost of the 11th unit.
- (b) A salesman is entitled to bonus in a year if his average quarterly sales are at least Rs 40,000. If his average sales of the first three quarters is Rs 35,000, find his minimum level of sales in the fourth quarter so that he becomes eligible for bonus.

Hint: See example 10.

30. (a) The monthly salaries of five persons were Rs 5,000, Rs 5,500, Rs 6,000, Rs 7,000 and Rs 20,000. Compute their mean salary. Would you regard this mean as typical of the salaries? Explain.
- (b) There are 100 workers in a company out of which 70 are males and 30 females. If a male worker earns Rs 100 per day and a female worker earns Rs. 70 per day, find average wage. Would you regard this as a typical wage? Explain

Hint: An average that is representative of most of the observations is said to be a typical average.

2.6 MEDIAN

Median of distribution is that value of the variate which divides it into two equal parts. In terms of frequency curve, the ordinate drawn at median divides the area under the curve into two equal parts. Median is a positional average because its value depends upon the

Determination of Median

(a) When individual observations are given

The following steps are involved in the determination of median :

- (i) The given observations are arranged in either ascending or descending order of magnitude.
- (ii) Given that there are n observations, the median is given by:

1. The size of $\left(\frac{n+1}{2}\right)$ th observations, when n is odd.

2. The mean of the sizes of $\frac{n}{2}$ th and $\left(\frac{n+1}{2}\right)$ th observations, when n is even.

Example 16: Find median of the following observations :

20, 15, 25, 28, 18, 16, 30.

Solution: Writing the observations in ascending order, we get 15, 16, 18, 20, 25, 28, 30.

Since $n = 7$, i.e., odd, the median is the size of $\left(\frac{7+1}{2}\right)$ th, i.e., 4th observation.

Hence, median, denoted by $M_d = 20$.

Note: The same value of M_d will be obtained by arranging the observations in descending order of magnitude.

Example 17: Find median of the data : 245, 230, 265, 236, 220, 250.

Solution: Arranging these observations in ascending order of magnitude, we get

220, 230, 236, 245, 250, 265. Here $n = 6$, i.e., even.

\therefore Median will be arithmetic mean of the size of $\frac{6}{2}$ th, i.e., 3rd and $\left(\frac{6}{2}+1\right)$ th,

i.e., 4th observations. Hence $M_d = \frac{236+245}{2} = 240.5$.

Remarks: Consider the observations: 13, 16, 16, 17, 17, 18, 19, 21, 23. On the basis of the method given above, their median is 17.

According to the above definition of median, "half (i.e., 50%) of the observations should be below 17 and half of the observations should be above 17". Here we may note that only 3 observations are below 17 and 4 observations are above it and hence, the definition of median given above is some what ambiguous. In order to avoid this ambiguity, the median of a distribution may also be defined in the following way :

Median of a distribution is that value of the variate such that at least half of the observations are less than or equal to it and at least half of the observations are greater than or equal to it.

Based on this definition, we find that there are 5 observations which are less than or equal to 17 and there are 6 observations which are greater than or equal to 17. Since $n = 9$, the numbers 5 and 6 are both more than half, i.e., 4.5. Thus, median of the distribution is 17.

Further, if the number of observations is even and the two middle most observations are not equal, e.g., if the observations are 2, 2, 5, 6, 7, 8, then there are 3 observations

$\left(\frac{n}{2} = 3\right)$ which are less than or equal to 5 and there are 4 (i.e., more than half) observations

which are greater than or equal to 5. Further, there are 4 observations which are less than or equal to 6 and there are 3 observations which are greater than or equal to 6. Hence, both 5 and 6 satisfy the conditions of the new definition of median. In such a case, any value lying in the closed interval [5, 6] can be taken as median. By convention

we take the middle value of the interval as median. Thus, median is $\frac{5+6}{2} = 5.5$

(b) When ungrouped frequency distribution is given

In this case, the data are already arranged in the order of magnitude. Here, cumulative frequency is computed and the median is determined in a manner similar to that of individual observations.

Example 18: Locate median of the following frequency distribution :

Variable (X) :	10	11	12	13	14	15	16
Frequency (f) :	8	15	25	20	12	10	5

Solution:

X :	10	11	12	13	14	15	16
f :	8	15	25	20	12	10	5
c.f. :	8	23	48	68	80	90	95

Here $N = 95$, which is odd. Thus, median is size of $\left[\frac{95+1}{2}\right]^{\text{th}}$
i.e., 48th observation. From the table 48th observation is 12, $\therefore M_d = 12$.

Alternative Method: $\frac{N}{2} = \frac{95}{2} = 47.5$ Looking at the frequency distribution we note that there are 48 observations which are less than or equal to 12 and there are 72 (i.e., $95 - 23$) observations which are greater than or equal to 12. Hence, median is 12.

Example 19: Locate median of the following frequency distribution :

X :	0	1	2	3	4	5	6	7
f :	7	14	18	36	51	54	52	20

Solution:

X	0	1	2	3	4	5	6	7
f	7	14	18	36	51	54	52	20
c.f.	7	21	39	75	126	180	232	252

Here $N = 252$, i.e., even.

Now $\frac{N}{2} = \frac{252}{2} = 126$ and $\frac{N}{2} + 1 = 127$.

\therefore Median is the mean of the size of 126th and 127th observation. From the table we note that 126th observation is 4 and 127th observation is 5.

$$\therefore M_d = \frac{4+5}{2} = 4.5$$

Alternative Method: Looking at the frequency distribution we note that there are 126 observations which are less than or equal to 4 and there are $252 - 75 = 177$ observations which are greater than or equal to 4. Similarly, observation 5 also satisfies this criterion.

Therefore, median = $\frac{4+5}{2} = 4.5$.

(c) When grouped frequency distribution is given (Interpolation formula)

Measures of Central Tendency

The determination of median, in this case, will be explained with the help of the following example.

Example 20: Suppose we wish to find the median of the following frequency distribution.

Classes	: 0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
Frequency	: 5	12	14	18	13	8

Solution: The median of a distribution is that value of the variate which divides the distribution into two equal parts. In case of a grouped frequency distribution, this implies that the ordinate drawn at the median divides the area under the histogram into two equal parts. Writing the given data in a tabular form, we have :

Classes (1)	Frequency (f) (2)	'Less than' type c. f. (3)	Frequency Density (4)
0 - 10	5	5	0.5
10 - 20	12	17	1.2
20 - 30	14	31	1.4
30 - 40	18	49	1.8
40 - 50	13	62	1.3
50 - 60	8	70	0.8

$$\left(\text{Note : frequency density in a class} = \frac{\text{frequency of the class}}{\text{Width of the class}} = \frac{f}{h} \right)$$

For the location of median, we make a histogram with heights of different rectangles equal to frequency density of the corresponding class. Such a histogram is shown below:

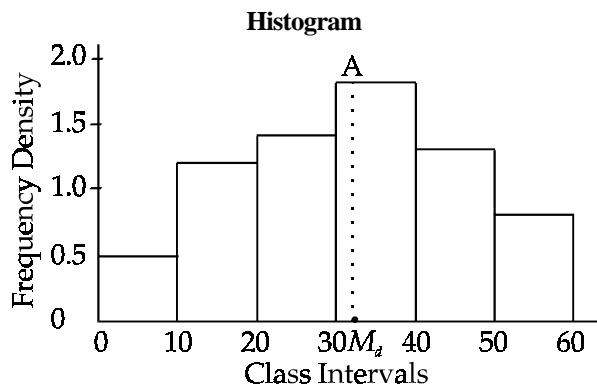


Figure : 2.1

Since the ordinate at median divides the total area under the histogram into two equal parts, therefore we have to find a point (like M_d as shown in the figure) on X - axis such that an ordinate (AM_d) drawn at it divides the total area under the histogram into two equal parts.

We may note here that area under each rectangle is equal to the frequency of the corresponding class.

$$\text{Since area} = \text{length} \times \text{breadth} = \text{frequency density} \times \text{width of class} = \frac{f}{h} \times h = f.$$

Thus, the total area under the histogram is equal to total frequency N. In the given example $N = 70$, therefore $\frac{N}{2} = 35$. We note that area of first three rectangles is

$5 + 12 + 14 = 31$ and the area of first four rectangles is $5 + 12 + 14 + 18 = 49$. Thus, median lies in the fourth class interval which is also termed as median class. Let the point, in median class, at which median lies be denoted by M_d . The position of this point

should be such that the ordinate M_d (in the above histogram) divides the area of median rectangle so that there are only $35 - 31 = 4$ observations to its left. From the histogram, we can also say that the position of M_d should be such that

$$\frac{M_d - 30}{40 - 30} = \frac{4}{18} \quad \dots (1)$$

$$\text{Thus, } M_d = \frac{40}{18} + 30 = 32.2$$

Writing the above equation in general notations, we have

$$\frac{M_d - L_m}{h} = \frac{\frac{N}{2} - C}{f_m} \text{ or } M_d = L_m + \left(\frac{\frac{N}{2} - C}{f_m} \right) h \quad \dots (2)$$

Where, L_m is lower limit, h is the width and f_m is frequency of the median class and C is the cumulative frequency of classes preceding median class. Equation (2) gives the required formula for the computation of median.

Remarks:

1. Since the variable, in a grouped frequency distribution, is assumed to be continuous we always take exact value of $\frac{N}{2}$, including figures after decimals, when N is odd.
2. The above formula is also applicable when classes are of unequal width.
3. Median can be computed even if there are open end classes because here we need to know only the frequencies of classes preceding or following the median class.

Determination of Median When 'greater than' type cumulative frequencies are given

By looking at the histogram, we note that one has to find a point denoted by M_d such that area to the right of the ordinate at M_d is 35. The area of the last two rectangles is $13 + 8 = 21$. Therefore, we have to get $35 - 21 = 14$ units of area from the median rectangle towards right of the ordinate. Let U_m be the upper limit of the median class. Then the formula for median in this case can be written as

$$\frac{U_m - M_d}{h} = \frac{\frac{N}{2} - C}{f_m} \text{ or } M_d = U_m - \frac{\frac{N}{2} - C}{f_m} \times h \quad \dots (3)$$

Note that C denotes the 'greater than' type cumulative frequency of classes following the median class. Applying this formula to the above example, we get

$$M_d = 40 - \frac{(35 - 21)}{18} \times 10 = 32.2$$

Example 21: Calculate median of the following data :

Height in inches :	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8	8 - 9	9 - 10	10 - 11
No. of saplings :	3	7	12	16	22	20	13	7

Calculation of Median

Class Intervals	Frequency (f)	'Less than' type c. f.
3 - 4	3	3
4 - 5	7	10
5 - 6	12	22
6 - 7	16	38
7 - 8	22	60
8 - 9	20	80
9 - 10	13	93
10 - 11	7	100

Since $\frac{N}{2} = \frac{100}{2} = 50$, the median class is 7 - 8. Further, $L_m = 7$, $h = 1$, $f_m = 22$ and $C = 38$.

Thus, $M_d = 7 + \frac{50-38}{22} \times 1 = 7.55$ inches.

Example 22: The following table gives the distribution of marks by 500 students in an examination. Obtain median of the given data.

Marks : 0 - 9 10 - 19 20 - 29 30 - 39 40 - 49 50 - 59 60 - 69 70 - 79

No. of Students : 30 40 50 48 24 162 132 14

Solution: Since the class intervals are inclusive, therefore, it is necessary to convert them into class boundaries.

Class Intervals	Class Boundaries	Frequency	'Less than' type c.f.
0-9	-0.5-9.5	30	30
10-19	9.5-19.5	40	70
20-29	19.5-29.5	50	120
30-39	29.5-39.5	48	168
40-49	39.5-49.5	24	192
50-59	49.5-59.5	162	354
60-69	59.5-69.5	132	486
70-79	69.5-79.5	14	500

Since $\frac{N}{2} = \frac{500}{2} = 250$, the median class is 49.5 - 59.5 and, therefore, $L_m = 49.5$, $h = 10$, $f_m = 162$, $C = 192$.

Thus, $M_d = 49.5 + \frac{250-192}{162} \times 10 = 53.08$ marks.

Example 23: The weekly wages of 1,000 workers of a factory are shown in the following table. Calculate median.

Weekly Wages (less than) : 425 475 525 575 625 675 725 775 825 875

No. of Workers : 2 10 43 123 293 506 719 864 955 1000

Solution: The above is a 'less than' type frequency distribution. This will first be converted into class intervals.

Class Intervals	Frequency	Less than c. f.
less than 425	2	2
425 - 475	8	10
475 - 525	33	43
525 - 575	80	123
575 - 625	170	293
625 - 675	213	506
675 - 725	213	719
725 - 775	145	864
775 - 825	91	955
825 - 875	45	1000

Since $\frac{N}{2} = 500$, the median class is 625 - 675. On substituting various values in the formula for median, we get

$$M_d = 625 + \frac{500 - 293}{213} \times 50 = \text{Rs } 673.59$$

Example 24: Find the median of the following data :

Age greater than (in yrs) : 0	10	20	30	40	50	60	70
No. of Persons : 230	218	200	165	123	73	28	8

Solution: Note that it is 'greater than' type frequency distribution.

Class Intervals	Greater than c.f.	Frequency
0 - 10	230	12
10 - 12	218	18
20 - 30	200	35
30 - 40	165	42
40 - 50	123	50
50 - 60	73	45
60 - 70	28	20
70 and above	8	8

Since $\frac{N}{2} = \frac{230}{2} = 115$, the median class is 40 - 50.

$$\text{Using the formula, } M_d = U_m - \frac{\frac{N}{2} - C}{f_m} \times h$$

$$= 50 - \frac{115 - 73}{50} \times 10 = 41.6 \text{ years}$$

Example 25: The following table gives the daily profits (in Rs) of 195 shops of a town. Calculate mean and median.

Profits : 50 - 60	60 - 70	70 - 80	80 - 90	90 - 100	100 - 110	110 - 120	120 - 130	130 - 140
No. of shops: 15	20	32	35	33	22	20	10	8

Solution: The calculations of \bar{X} and M_d are shown below:

Class Intervals	f	Mid-value (X)	$u = \frac{X - 95}{10}$	fu	Less than c.f.
50-60	15	55	-4	-60	15
60-70	20	65	-3	-60	35
70-80	32	75	-2	-64	67
80-90	35	85	-1	-35	102
90-100	33	95	0	0	135
100-110	22	105	1	22	157
110-120	20	115	2	40	177
120-130	10	125	3	30	187
130-140	8	135	4	32	195
Total	195			-95	

$$\bar{X} = A + \frac{\sum fu}{N} \times h = 95 - \frac{95}{195} \times 10 = \text{Rs } 90.13$$

Since $\frac{N}{2} = \frac{195}{2} = 97.5$, the median class is 80 - 90.

$$\therefore Md = 80 + \frac{97.5 - 67}{35} \times 10 = \text{Rs } 88.71$$

Example 26: Find median of the following distribution :

Mid - Values	: 1500 2500 3500 4500 5500 6500 7500
Frequency	: 27 32 65 78 58 32 8

Solution: Since the mid-values are equally spaced, the difference between their two successive values will be the width of each class interval. This width is 1,000. On subtracting and adding half of this, i.e., 500 to each of the mid-values, we get the lower and the upper limits of the respective class intervals. After this, the calculation of median can be done in the usual way.

Mid - Values	Class Intervals	Frequency	c. f. (less than)
1500	1000 - 2000	27	27
2500	2000 - 3000	32	59
3500	3000 - 4000	65	124
4500	4000 - 5000	78	202
5500	5000 - 6000	58	260
6500	6000 - 7000	32	292
7500	7000 - 8000	8	300

Since $\frac{N}{2} = 150$, the median class is 4,000 - 5,000.

$$\text{Hence } M_d = 4,000 + \frac{150 - 124}{78} \times 1,000 = 4,333.33.$$

Determination of Missing Frequencies

If the frequencies of some classes are missing, however, the median of the distribution is known, then these frequencies can be determined by the use of median formula.

Example 27: The following table gives the distribution of daily wages of 900 workers. However, the frequencies of the classes 40 - 50 and 60 - 70 are missing. If the median of the distribution is Rs 59.25, find the missing frequencies.

Wages (Rs)	:	30-40	40-50	50-60	60-70	70-80
No. of Workers	:	120	?	200	?	185

Solution: Let f_1 and f_2 be the frequencies of the classes 40 - 50 and 60 - 70 respectively.

Class Intervals	Frequency	c.f. (less than)
30-40	120	120
40-50	f_1	$120 + f_1$
50-60	200	$320 + f_1$
60-70	f_2	$320 + f_1 + f_2$
70-80	185	900

Since median is given as 59.25, the median class is 50 - 60.

Therefore, we can write

$$59.25 = 50 + \frac{450 - (120 + f_1)}{200} \times 10 = 50 + \frac{330 - f_1}{20}$$

$$\text{or } 9.25 \times 20 = 330 - f_1 \text{ or } f_1 = 330 - 185 = 145$$

$$\text{Further, } f_2 = 900 - (120 + 145 + 200 + 185) = 250.$$

Graphical location of Median

So far we have calculated median by the use of a formula. Alternatively, it can be determined graphically, as illustrated in the following example.

Example 28: The following table shows the daily sales of 230 footpath sellers of Chandni Chowk :

Sales (in Rs) :	0 - 500	500 - 1000	1000 - 1500	1500 - 2000
No. of Sellers :	12	18	35	42
Sales (in Rs) :	2000 - 2500	2500 - 3000	3000 - 3500	3500 - 4000
No. of Sellers :	50	45	20	8

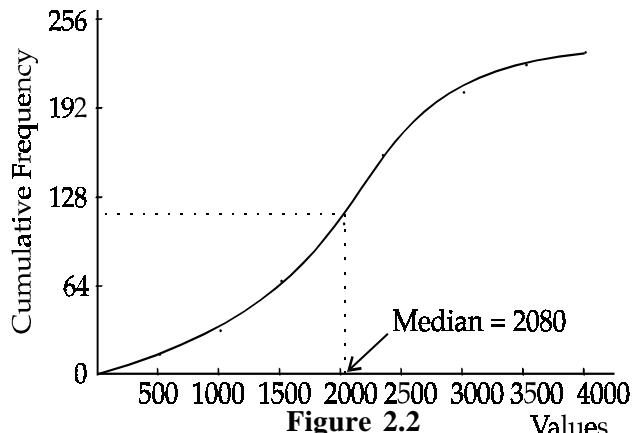
Locate the median of the above data using

- (i) only the less than type ogive, and
- (ii) both, the less than and the greater than type ogives.

Solution: To draw ogives, we need to have a cumulative frequency distribution.

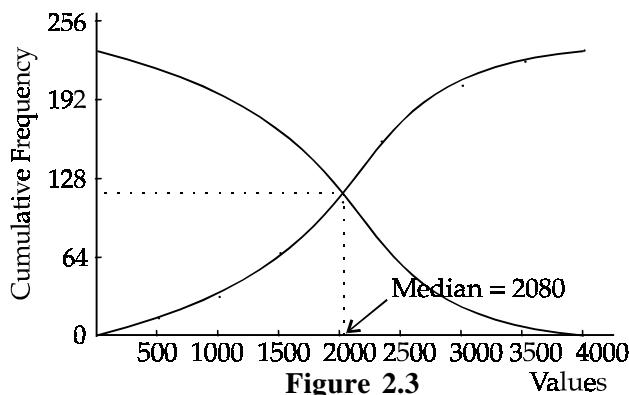
Class Intervals	Frequency	Less than c. f.	More than c. f.
0 - 500	12	12	230
500 - 1000	18	30	218
1000 - 1500	35	65	200
1500 - 2000	42	107	165
2000 - 2500	50	157	123
2500 - 3000	45	202	73
3000 - 3500	20	222	28
3500 - 4000	8	230	8

(i) *Using the less than type ogive*



The value $\frac{N}{2} = 115$ is marked on the vertical axis and a horizontal line is drawn from this point to meet the ogive at point S. Drop a perpendicular from S. The point at which this meets X- axis is the median.

(ii) *Using both types of ogives*



A perpendicular is dropped from the point of intersection of the two ogives. The point at which it intersects the X-axis gives median. It is obvious from Fig. 2.2 and 2.3 that median = 2080.

Properties of Median

1. It is a positional average.
2. It can be shown that the sum of absolute deviations is minimum when taken from median. This property implies that median is centrally located.

Merits and Demerits of Median

(a) Merits

1. It is easy to understand and easy to calculate, especially in series of individual observations and ungrouped frequency distributions. In such cases it can even be located by inspection.
2. Median can be determined even when class intervals have open ends or not of equal width.
3. It is not much affected by extreme observations. It is also independent of range or dispersion of the data.
4. Median can also be located graphically.
5. It is centrally located measure of average since the sum of absolute deviation is minimum when taken from median.
6. It is the only suitable average when data are qualitative and it is possible to rank various items according to qualitative characteristics.
7. Median conveys the idea of a typical observation.

(b) Demerits

1. In case of individual observations, the process of location of median requires their arrangement in the order of magnitude which may be a cumbersome task, particularly when the number of observations is very large.
2. It, being a positional average, is not capable of being treated algebraically.
3. In case of individual observations, when the number of observations is even, the median is estimated by taking mean of the two middle-most observations, which is not an actual observation of the given data.
4. It is not based on the magnitudes of all the observations. There may be a situation where different sets of observations give same value of median. For example, the following two different sets of observations, have median equal to 30.

Set I : 10, 20, 30, 40, 50 and Set II : 15, 25, 30, 60, 90.

5. In comparison to arithmetic mean, it is much affected by the fluctuations of sampling.
6. The formula for the computation of median, in case of grouped frequency distribution, is based on the assumption that the observations in the median class are uniformly distributed. This assumption is rarely met in practice.
7. Since it is not possible to define weighted median like weighted arithmetic mean, this average is not suitable when different items are of unequal importance.

Uses

1. It is an appropriate measure of central tendency when the characteristics are not measurable but different items are capable of being ranked.
2. Median is used to convey the idea of a typical observation of the given data.
3. Median is the most suitable measure of central tendency when the frequency distribution is skewed. For example, income distribution of the people is generally positively skewed and median is the most suitable measure of average in this case.
4. Median is often computed when quick estimates of average are desired.
5. When the given data has class intervals with open ends, median is preferred as a measure of central tendency since it is not possible to calculate mean in this case.

Check Your Progress 2.1

1. What are the merits and demerits of Mean and Median?
2. Find Arithmetic mean of first ten prime numbers.

Notes: (a) Write your answer in the space given below.
 (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

2.7 OTHER PARTITION OR POSITIONAL MEASURES

Median of a distribution divides it into two equal parts. It is also possible to divide it into more than two equal parts. The values that divide a distribution into more than two equal parts are commonly known as partition values or fractiles. Some important partition values are discussed in the following sections.

Quartiles

The values of a variable that divide a distribution into four equal parts are called quartiles. Since three values are needed to divide a distribution into four parts, there are three quartiles, viz. Q_1 , Q_2 and Q_3 , known as the first, second and the third quartile respectively.

For a discrete distribution, the first quartile (Q_1) is defined as that value of the variate such that at least 25% of the observations are less than or equal to it and at least 75% of the observations are greater than or equal to it.

For a continuous or grouped frequency distribution, Q_1 is that value of the variate such that the area under the histogram to the left of the ordinate at Q_1 is 25% and the area to its right is 75%. The formula for the computation of Q_1 can be written by making suitable changes in the formula of median.

After locating the first quartile class, the formula for Q_1 can be written as follows:

$$Q_1 = L_{Q_1} + \frac{\left(\frac{N}{4} - C\right)}{f_{Q_1}} \times h$$

Here, L_{Q_1} is lower limit of the first quartile class, h is its width, f_{Q_1} is its frequency and C is cumulative frequency of classes preceding the first quartile class.

By definition, the second quartile is median of the distribution. The third quartile (Q_3) of a distribution can also be defined in a similar manner.

For a discrete distribution, Q_3 is that value of the variate such that at least 75% of the observations are less than or equal to it and at least 25% of the observations are greater than or equal to it.

For a grouped frequency distribution, Q_3 is that value of the variate such that area under the histogram to the left of the ordinate at Q_3 is 75% and the area to its right is 25%. The formula for computation of Q_3 can be written as

$$Q_3 = L_{Q_3} + \frac{\left(\frac{3N}{4} - C\right)}{f_{Q_3}} \times h, \text{ where the symbols have their usual meaning.}$$

Deciles divide a distribution into 10 equal parts and there are, in all, 9 deciles denoted as D_1, D_2, \dots, D_9 respectively.

For a discrete distribution, the i th decile D_i is that value of the variate such that at least $(10i)\%$ of the observations are less than or equal to it and at least $(100 - 10i)\%$ of the observations are greater than or equal to it ($i = 1, 2, \dots, 9$).

For a continuous or grouped frequency distribution, D_i is that value of the variate such that the area under the histogram to the left of the ordinate at D_i is $(10i)\%$ and the area to its right is $(100 - 10i)\%$. The formula for the i th decile can be written as

$$D_i = L_{D_i} + \frac{\left(\frac{iN}{10} - C\right)}{f_{D_i}} \times h \quad (i = 1, 2, \dots, 9)$$

Percentiles

Percentiles divide a distribution into 100 equal parts and there are, in all, 99 percentiles denoted as $P_1, P_2, \dots, P_{25}, \dots, P_{40}, \dots, P_{60}, \dots, P_{99}$ respectively.

For a discrete distribution, the k th percentile P_k is that value of the variate such that at least $k\%$ of the observations are less than or equal to it and at least $(100 - k)\%$ of the observations are greater than or equal to it.

For a grouped frequency distribution, P_k is that value of the variate such that the area under the histogram to the left of the ordinate at P_k is $k\%$ and the area to its right is $(100 - k)\%$. The formula for the k th percentile can be written as

$$P_k = L_{P_k} + \frac{\left(\frac{kN}{100} - C\right)}{f_{P_k}} \times h, \quad (k = 1, 2, \dots, 99)$$

Remarks :

- (i) We may note here that $P_{25} = Q_1, P_{50} = D_5 = Q_2 = M_d, P_{75} = Q_3, P_{10} = D_1, P_{20} = D_2$, etc.
- (ii) In continuation of the above, the partition values are known as Quintiles (Octiles) if a distribution is divided into 5 (8) equal parts.
- (iii) The formulae for various partition values of a grouped frequency distribution, given so far, are based on 'less than' type cumulative frequencies. The corresponding formulae based on 'greater than' type cumulative frequencies can be written in a similar manner, as given below:

$$Q_1 = U_{Q_1} - \frac{\left(\frac{3N}{4} - C\right)}{f_{Q_1}} \times h, \quad Q_3 = U_{Q_3} - \frac{\left(\frac{N}{4} - C\right)}{f_{Q_3}} \times h$$

$$D_i = U_{D_i} - \frac{\left[\left(N - \frac{iN}{10}\right) - C\right]}{f_{D_i}} \times h, \quad P_k = U_{P_k} - \frac{\left[\left(N - \frac{kN}{100}\right) - C\right]}{f_{P_k}} \times h$$

Here $U_{Q_1}, U_{Q_3}, U_{D_i}, U_{P_k}$ are the upper limits of the corresponding classes and C denotes the greater than type cumulative frequencies.

Example 29: Locate Median, $Q_1, Q_3, D_4, D_7, P_{15}, P_{60}$ and P_{90} from the following data :

Daily Profit (in Rs) :	75	76	77	78	79	80	81	82	83	84	85
No. of Shops :	15	20	32	35	33	22	20	10	8	3	2

Solution: First we calculate the cumulative frequencies, as in the following table :

Daily Profit (in Rs)	75	76	77	78	79	80	81	82	83	84	85
No. of Shops (f)	15	20	32	35	33	22	20	10	8	3	2
Less than c.f.	15	35	67	102	135	157	177	187	195	198	200

1. **Determination of Median:** Here $\frac{N}{2} = 100$. From the cumulative frequency column, we note that there are 102 (greater than 50% of the total) observations that are less than or equal to 78 and there are 133 observations that are greater than or equal to 78. Therefore, $M_d = \text{Rs } 78$.
2. **Determination of Q_1 and Q_3 :** First we determine $\frac{N}{4}$ which is equal to 50. From the cumulative frequency column, we note that there are 67 (which is greater than 25% of the total) observations that are less than or equal to 77 and there are 165 (which is greater than 75% of the total) observations that are greater than or equal to 77. Therefore, $Q_1 = \text{Rs } 77$. Similarly, $Q_3 = \text{Rs } 80$.
3. **Determination of D_4 and D_7 :** From the cumulative frequency column, we note that there are 102 (greater than 40% of the total) observations that are less than or equal to 78 and there are 133 (greater than 60% of the total) observations that are greater than or equal to 78. Therefore, $D_4 = \text{Rs } 78$. Similarly, $D_7 = \text{Rs } 80$.
4. **Determination of P_{15} , P_{60} and P_{90} :** From the cumulative frequency column, we note that there are 35 (greater than 15% of the total) observations that are less than or equal to 76 and there are 185 (greater than 85% of the total) observations that are greater than or equal to 76. Therefore, $P_{15} = \text{Rs } 76$. Similarly, $P_{60} = \text{Rs } 79$ and $P_{90} = \text{Rs } 82$.

Example 30: Calculate median, quartiles, 3rd and 6th deciles and 40th and 70th percentiles, from the following data:

Wages per Week (in Rs)	:	50-100	100-150	150-200	200-250	250-300
No. of Workers	:	15	40	35	60	125
Wages per Week (in Rs)	:	300-350	350-400	400-450	450-500	
No. of Workers	:	100	70	40	15	

Also determine (i) The percentage of workers getting weekly wages between Rs 125 and Rs 260 and (ii) percentage of worker getting wages greater than Rs 340.

Solution: First we make a cumulative frequency distribution table :

Class Intervals	Frequency	c. f. (less than)
50 - 100	15	15
100 - 150	40	55
150 - 200	35	90
200 - 250	60	150
250 - 300	125	275
300 - 350	100	375
350 - 400	70	445
400 - 450	40	485
450 - 500	15	500

- (i) **Calculation of median:** Here $N = 500$ so that $\frac{N}{2} = 250$. Thus, median class is 250 - 300 and hence $L_m = 250$, $f_m = 125$, $h = 50$ and $C = 150$.

Substituting these values in the formula for median, we get

$$Md = 250 + \frac{250-150}{125} \times 50 = \text{Rs } 290$$

(ii) ***Calculation of Quartiles:***

Measures of Central Tendency

- (a) For Q_1 , we first find $\frac{N}{4}$ which is equal to 125. The first quartile class is 200 - 250 and hence $L_{Q_1} = 200$, $f_{Q_1} = 60$, $h = 50$ and $C = 90$.

$$\therefore Q_1 = 200 + \frac{125 - 90}{60} \times 50 = \text{Rs } 229.17$$

- (b) For Q_3 we first find $\frac{3N}{4}$ which is equal to 375. The third quartile class is 300 - 350 and hence $L_{Q_3} = 300$, $f_{Q_3} = 100$, $h = 50$ and $C = 275$.

$$\therefore Q_3 = 300 + \frac{375 - 275}{100} \times 50 = \text{Rs } 350$$

(iii) ***Calculation of Deciles:***

- (a) For D_3 we first find $\frac{3N}{10}$ which is equal to 150. The third decile class is 200 - 250 and hence $L_{D_3} = 200$, $f_{D_3} = 60$, $h = 50$, $C = 90$.

$$\therefore D_3 = 200 + \frac{150 - 90}{60} \times 50 = \text{Rs } 250$$

- (b) For D_6 we first find $\frac{6N}{10}$ which is equal to 300. The sixth decile class is 300 - 350 and hence $L_{D_6} = 300$, $f_{D_6} = 100$, $h = 50$ and $C = 275$.

$$\therefore D_6 = 300 + \frac{300 - 275}{100} \times 50 = \text{Rs } 312.50$$

(iv) ***Calculation of percentiles:***

- (a) For P_{40} we first find $\frac{40N}{100}$ which is equal to 200. The 40th percentile class is 250 - 300 and hence $L_{P_{40}} = 250$, $f_{P_{40}} = 125$, $h = 50$ and $C = 150$.

$$\therefore P_{40} = 250 + \frac{200 - 150}{125} \times 50 = \text{Rs } 270$$

- (b) For P_{70} we first find $\frac{70N}{100}$ which is equal to 350. The 70th percentile class is 300 - 350 and hence $L_{P_{70}} = 300$, $f_{P_{70}} = 100$, $h = 50$ and $C = 275$.

$$\therefore P_{70} = 300 + \frac{350 - 275}{100} \times 50 = \text{Rs } 337.5$$

(v) ***Determination of percentage of workers getting wages between Rs 125 and Rs 260:***

Let x be the percentage of workers getting wage less than 125. Since 125 lies in the class 100 - 150, this is x th percentiles class. Using the formula for x th percentile we have

$$125 = 100 + \frac{\frac{x}{100} \times 50 - 15}{40} \quad \text{or} \quad 5x - 15 = 20 \Rightarrow x = 7$$

Further, let y be the percentage of workers getting wages less than 260. Since 260 lies in the class 250 - 300, this is y th percentile class. Using the relevant formula, we have

$$260 = 250 + \frac{5y - 150}{125} \times 50 \text{ or } 5y - 150 = 25 \text{ or } y = 35$$

Hence percentage of workers getting wages between Rs 125 and Rs 260 is given by $35 - 7 = 28\%$.

Alternative Method

The number of workers getting wages between 125 and 260 can be written directly as

$$= \frac{150 - 125}{50} \times 40 + 35 + 60 + \frac{260 - 250}{50} \times 125 = 20 + 35 + 60 + 25 = 140.$$

$$\therefore \text{Percentage of workers} = \frac{140}{500} \times 100 = 28\%.$$

(vi) **Determination of percentage of workers getting wages greater than Rs 340:**

Since we have already computed 'less than' type cumulative frequencies, in the above table, we shall first find percentage of workers getting wages less than 340. Let x be this percentage. Also x th percentiles class is 300 - 350.

$$\therefore 340 = 300 + \frac{5x - 275}{100} \times 50 \text{ or } 5x - 275 = 80 \text{ or } x = 71$$

Hence, percentage of workers getting wages greater than Rs 340 is $(100 - 71) = 29\%$.

Alternative Method

This percentage can also be obtained directly as shown below.

The percentage of workers getting wages greater than Rs 340

$$= \frac{350 - 340}{50} \times 100 + 70 + 40 + 15 = 145$$

$$\therefore \text{Percentage} = \frac{145}{500} \times 100 = 29\%$$

Example 31: From the following table, showing the wage distribution of workers, find

- (i) the range of incomes earned by middle 50% of the workers,
- (ii) the range of incomes earned by middle 80% of the workers,
- (iii) the percentage of workers earning between Rs 550 and Rs 880.

Monthly Income (Rs)	No. of Workers
0-200	150
0-400	250
0-600	330
0-800	380
0-1000	400

Solution: The above table gives a 'less than' type cumulative frequency distribution. Therefore, we can rewrite the above table as :

Monthly Income (Rs)	c.f. (less than)	Frequency (f)
0-200	150	150
200-400	250	100
400-600	330	80
600-800	380	50
800-1000	400	20

- (i) The range of incomes earned by middle 50% of the workers is given by $Q_3 - Q_1$.

Measures of Central Tendency

$$\text{Now } Q_1 = 0 + \frac{100}{150} \times 200 = \text{Rs } 133.33$$

$$\text{and } Q_3 = 400 + \frac{300-250}{80} \times 200 = \text{Rs } 525.$$

$$\text{Thus, } Q_3 - Q_1 = 525 - 133.33 = \text{Rs. } 391.67.$$

- (ii) The range of incomes of middle 80% of the workers is given by $P_{90} - P_{10}$.

$$\text{Now } P_{10} = 0 + \frac{\frac{10 \times 400}{100} - 0}{150} \times 200 = \text{Rs } 53.33$$

$$\text{and } P_{90} = 600 + \frac{\frac{90 \times 400}{100} - 330}{50} \times 200 = \text{Rs } 720.$$

$$\text{Thus, } P_{90} - P_{10} = 720 - 53.33 = \text{Rs } 666.67.$$

- (iii) The No. of workers earning between Rs 550 and Rs 880 is given by

$$\frac{600-550}{200} \times 80 + 50 + \frac{880-800}{200} \times 20 = 78.$$

$$\therefore \text{Percentage of workers} = \frac{78}{400} \times 100 = 19.5\%$$

Example 32: The following incomplete table gives the number of students in different age groups of a town. If the median of the distribution is 11 years, find out the missing frequencies.

Age Group	: 0 - 5	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30	Total
No. of Students	15	125	?	66	?	4	300

Solution: Let x be the frequency of age group 10 - 15. Then the frequency of the age group 20 - 25 will be $300 - (15 + 125 + x + 66 + 4) = 90 - x$.

Making a cumulative frequency table we have

Age Groups	No. of Students	c.f. (less than)
0-5	15	15
5-10	125	140
10-15	x	$140+x$
15-20	66	$206+x$
20-25	$90-x$	296
25-30	4	300

Here $\frac{N}{2} = \frac{300}{2} = 150$. Since median is given as 11, the median class is 10 - 15.

$$\text{Hence, } 11 = 10 + \frac{150-140}{x} \times 5 \text{ or } x = 50.$$

Also, frequency of the age group 20 - 25 is $90 - 50 = 40$.

Exercise with Hints

1. The following table gives weekly income of 24 families in a certain locality :

S. No. of family	1	2	3	4	5	6	7	8	9	10	11	12
Weekly Income	60	400	86	95	100	150	110	74	90	92	280	180
S. No. of family	13	14	15	16	17	18	19	20	21	22	23	24
Weekly Income	96	98	104	75	80	94	100	75	600	82	200	84

Calculate M_d , Q_1 , Q_3 , D_4 , D_7 , P_{20} , P_{45} , and P_{95} .

Hint: Arrange the data in ascending or descending order of magnitude and then calculate various values. For calculation of Q_1 there are two values satisfying the definition. These two values are 82 and 84. Thus, Q_1 can be any value in the closed interval [82, 84]. By convention, the mid-value of the interval is taken as Q_1 .

2. Calculate the value of M_d , Q_1 , Q_3 , D_2 , D_8 , P_{35} , P_{48} , and P_{68} , from the following data:

Classes	: below	10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
Frequency	:	1	2	5	7	10	7	5	2	1

Hint: See example 30.

3. Find median from the following data:

Wages/Week (Rs)	: 50-59	60-69	70-79	80-89	90-99	100-109	110-119
No. of Workers	: 15	40	50	60	45	40	15

Hint: This is a distribution with inclusive class intervals. To compute median, these are to be converted into exclusive intervals like 49.5 - 59.5, 59.5 - 69.5, etc.

4. The following table gives the distribution of wages of 65 employees in a factory :

Wages (\geq)	: 50	60	70	80	90	100	110	120
No. of employees	: 65	57	47	31	17	7	2	0

Draw a 'less than type' ogive from the above data and estimate the number of employees earning at least Rs 63 but less than Rs 75.

Hint: To draw a 'less than' type ogive, the distribution is to be converted into 'less than' type cumulative frequencies.

5. The following table shows the age distribution of persons in a particular region:

Age (years)	No. of Persons ('000)
Below 10	2
Below 20	5
Below 30	9
Below 40	12
Below 50	14
Below 60	15
Below 70	15.5
70 and above	15.6

- (i) Find median age.
- (ii) Why is the median a more suitable measure of central tendency than mean in this case?

Hint: Median is suitable here because the upper limit of the last class is not known and therefore, mean cannot be satisfactorily calculated.

6. A number of particular articles have been classified according to their weights. After drying for two weeks the same articles have again been weighed and similarly classified. It is known that median weight in the first weighing was 20.83 oz. while in second weighing it was 17.35 oz. Some frequencies a and b in the first weighing and x and y in the second weighing, are missing. It is known that $a = \frac{1}{3}x$ and

$$b = \frac{1}{2}y. \text{ Find the missing frequencies.}$$

Classes		0 - 5	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30
Frequency	1st weighing	a	b	11	52	75	22
	2nd weighing	x	y	40	50	30	28

$$\text{Hint: } 20.83 = 20 + \frac{\left[\left(\frac{160+a+b}{2} \right) - (63+a+b) \right]}{75} \times 5 \quad \dots (1)$$

$$\text{and } 17.35 = 15 + \frac{\left[\left(\frac{148+x+4}{2} \right) - (40+x+y) \right]}{50} \times 5 \quad \dots (2)$$

Put $x = 3a$ and $y = 2b$ in equation (2) and solve (1) and (2) simultaneously.

7. The percentage distribution of regularly employed workers who commute between home and work place by foot and those who use cycles, according to the distance is given below. How will you find the mean distance and the median distance of the walkers and cyclists? State your assumptions carefully.

<i>Distance in kms</i>	<i>Walkers</i>	<i>Cyclists</i>
<i>less than 1/4</i>	45.3	1.1
1/4 - 1/2	21.1	6.0
1/2 - 1	15.2	9.6
1 - 2	9.8	17.9
2 - 3	5.3	20.5
3 - 4	2.2	19.2
4 - 5	0.6	19.2
<i>above 5</i>	0.5	10.5

- Hint:** The given percentage of walkers and cyclists can be taken as frequencies. For calculation of mean, the necessary assumption is that the width of the first class is equal to the width of the following class, i.e., $\frac{1}{4}$. On this assumption, the lower limit of the first class can be taken as 0. Similarly, on the assumption that width of the last class is equal to the width of last but one class, the upper limit of last class can be taken as 6. No assumption is needed for the calculation of median.

8. In a factory employing 3,000 persons, 5 percent earn less than Rs 3 per hour, 580 earn Rs 3.01 to 4.50 per hour, 30 percent earn from Rs 4.51 to Rs 6.00 per hour, 500 earn from 6.01 to Rs 7.5 per hour, 20 percent earn from Rs 7.51 to Rs 9.00 per hour and the rest earn Rs 9.01 or more per hour. What is the median wage?

- Hint:** Write down the above information in the form of a frequency distribution. The class intervals given above are inclusive type. These should be converted into exclusive type for the calculation of median.

9. The distribution of 2,000 houses of a new locality according to their distance from a milk booth is given in the following table :

<i>Distance (in metres)</i>	<i>No. of Houses</i>	<i>Distance (in metres)</i>	<i>No. of Houses</i>
0 - 50	20	350 - 400	275
50 - 100	30	400 - 450	400
100 - 150	35	450 - 500	325
150 - 200	46	500 - 550	205
200 - 250	50	550 - 600	184
250 - 300	105	600 - 650	75
300 - 350	200	650 - 700	50

- (i) Calculate the median distance of a house from milk booth.
- (ii) In second phase of the construction of the locality, 500 additional houses were constructed out of which the distances of 200, 150 and 150 houses from the milk booth were in the intervals 450 - 500, 550 - 600 and 650 - 700 meters respectively. Calculate the median distance, taking all the 2500 houses into account.

Hint: Add 200, 150 and 150 to the respective frequencies of the class intervals 450 - 500, 550 - 600 and 650 - 700.

10. The monthly salary distribution of 250 families in a certain locality of Agra is given below.

Monthly Salary	No. of Families	Monthly Salary	No. of Families
more than 0	250	more than 2000	55
more than 500	200	more than 2500	30
more than 1000	120	more than 3000	15
more than 1500	80	more than 3500	5

Draw a 'less than' ogive for the data given above and hence find out :

- (i) The limits of the income of the middle 50% of the families. (ii) If income tax is to be levied on families whose income exceeds Rs 1800 p.m., calculate the percentage of families which will be paying income tax.

Hint: See example 23.

11. The following table gives the frequency distribution of marks of 800 candidates in an examination :

Marks	: 0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
No. of candidates	: 10	40	80	140	170
Marks	: 50 - 60	60 - 70	70 - 80	80 - 90	90 - 100
No. of candidates	: 130	100	70	40	20

Draw 'less than' and 'more than' type ogives for the above data and answer the following from the graph :

- (i) If the minimum marks required for passing are 35, what percentage of candidates pass the examination?
(ii) It is decided to allow 80% of the candidate to pass, what should be the minimum marks for passing?
(iii) Find the median of the distribution.

Hint: See example 28.

12. Following are the marks obtained by a batch of 10 students in a certain class test in statistics (X) and accountancy (Y).

Roll No.	:	1	2	3	4	5	6	7	8	9	10
X	:	63	64	62	32	30	60	47	46	35	28
Y	:	68	66	35	42	26	85	44	80	33	72

In which subject the level of knowledge of student is higher?

Hint: Compare median of the two series.

13. The mean and median marks of the students of a class are 50% and 60% respectively. Is it correct to say that majority of the students have secured more than 50% marks? Explain.

Hint: It is given that at least 50% of the students have got 60% or more marks.

14. The monthly wages of 7 workers of a factory are : Rs 1,000, Rs 1,500, Rs 1,700, Rs 1,800, Rs 1,900, Rs 2,000 and Rs 3,000. Compute mean and median. Which measure is more appropriate? Which measure would you use to describe the situation if you were (i) a trade union leader, (ii) an employer?

Hint: (i) median, (ii) mean.

15. A boy saves Re. 1 on the first day, Rs 2 on the second day, Rs 31 on the 31st day of a particular month. Compute the mean and median of his savings per day. If his father contributes Rs 10 and Rs 100 on the 32nd and 33rd day respectively, compute mean and median of his savings per day. Comment upon the results.

Hint: Mean is too much affected by extreme observations.

Mode is that value of the variate which occurs maximum number of times in a distribution and around which other items are densely distributed. In the words of Croxton and Cowden, "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded the most typical of a series of values." Further, according to A.M. Tuttle, "Mode is the value which has the greatest frequency density in its immediate neighbourhood."

If the frequency distribution is regular, then mode is determined by the value corresponding to maximum frequency. There may be a situation where concentration of observations around a value having maximum frequency is less than the concentration of observations around some other value. In such a situation, mode cannot be determined by the use of maximum frequency criterion. Further, there may be concentration of observations around more than one value of the variable and, accordingly, the distribution is said to be bi-modal or multi-modal depending upon whether it is around two or more than two values.

The concept of mode, as a measure of central tendency, is preferable to mean and median when it is desired to know the most typical value, e.g., the most common size of shoes, the most common size of a ready-made garment, the most common size of income, the most common size of pocket expenditure of a college student, the most common size of a family in a locality, the most common duration of cure of viral-fever, the most popular candidate in an election, etc.

Determination of Mode

(a) *When data are either in the form of individual observations or in the form of ungrouped frequency distribution*

Given individual observations, these are first transformed into an ungrouped frequency distribution. The mode of an ungrouped frequency distribution can be determined in two ways, as given below :

- (i) By inspection or
- (ii) By method of Grouping
- (iii) **By inspection:** When a frequency distribution is fairly regular, then mode is often determined by inspection. It is that value of the variate for which frequency is maximum. By a fairly regular frequency distribution we mean that as the values of the variable increase the corresponding frequencies of these values first increase in a gradual manner and reach a peak at certain value and, finally, start declining gradually in, approximately, the same manner as in case of increase.

Example 33: Compute mode of the following data :

3, 4, 5, 10, 15, 3, 6, 7, 9, 12, 10, 16, 18,
20, 10, 9, 8, 19, 11, 14, 10, 13, 17, 9, 11

Solution: Writing this in the form of a frequency distribution, we get

Values	:	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Frequency	:	2	1	1	1	1	1	3	4	2	1	1	1	1	1	1	1	1	1

$$\therefore \text{Mode} = 10$$

Remarks :

- (i) If the frequency of each possible value of the variable is same, there is no mode.
- (ii) If there are two values having maximum frequency, the distribution is said to be bi-modal.

Example 34: Compute mode of the following distribution:

X:	5	10	15	20	25	30	35	40
f:	2	4	6	10	15	9	5	4

Solution: The given distribution is fairly regular. Therefore, the mode can be determined just by inspection. Since for X = 25 the frequency is maximum, mode = 25.

- (ii) **By method of Grouping:** This method is used when the frequency distribution is not regular. Let us consider the following example to illustrate this method.

Example 35: Determine the mode of the following distribution

X :	10	11	12	13	14	15	16	17	18	19
f :	8	15	20	100	98	95	90	75	50	30

Solution: This distribution is not regular because there is sudden increase in frequency from 20 to 100. Therefore, mode cannot be located by inspection and hence the method of grouping is used. Various steps involved in this method are as follows :

- (i) Prepare a table consisting of 6 columns in addition to a column for various values of X.
- (ii) In the first column, write the frequencies against various values of X as given in the question.
- (iii) In second column, the sum of frequencies, starting from the top and grouped in twos, are written.
- (iv) In third column, the sum of frequencies, starting from the second and grouped in twos, are written.
- (v) In fourth column, the sum of frequencies, starting from the top and grouped in threes are written.
- (vi) In fifth column, the sum of frequencies, starting from the second and grouped in threes are written.
- (vii) In the sixth column, the sum of frequencies, starting from the third and grouped in threes are written.

The highest frequency total in each of the six columns is identified and analysed to determine mode. We apply this method for determining mode of the above example.

X	f (1)	(2)	(3)	(4)	(5)	(6)
10	8					
11	15	23				
12	20		35			
13	100	120		43		
14	98		198		135	
15	95	193		293		218
16	90		185		283	
17	75	165		215		260
18	50		125		155	
19	30	80				

Analysis Table

Columns	V	A	R	I	A	B	L	E		
	10	11	12	13	14	15	16	17	18	19
1					1					
2						1	1			
3					1	1				
4					1	1	1			
5						1	1	1		
6							1	1	1	
Total	0	0	0	3	4	4	2	1	0	0

Since the value 14 and 15 are both repeated maximum number of times in the analysis table, therefore, mode is ill defined. Mode in this case can be approximately located by the use of the following formula, which will be discussed later, in this chapter.

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

Calculation of Median and Mean

X	10	11	12	13	14	15	16	17	18	19	Total
f	8	15	20	100	98	95	90	75	50	30	581
c. f.	8	23	43	143	241	336	426	501	551	581	
fX	80	165	240	1300	1372	1425	1440	1275	900	570	8767

$$\text{Median} = \text{Size of } \left(\frac{581+1}{2} \right) \text{ th, i.e., } 291\text{st observation} = 15. \text{ Mean} = \frac{8767}{581} = 15.09$$

$$\therefore \text{Mode} = 3 \times 15 - 2 \times 15.09 = 45 - 30.18 = 14.82$$

Remarks: If the most repeated values, in the above analysis table, were not adjacent, the distribution would have been bi-modal, i.e., having two modes

Example 36: From the following data regarding weights of 60 students of a class, find modal weight :

Weight	:	50	51	52	53	54	55	56	57	58	59	60
No. of Students	:	2	4	5	6	8	5	4	7	11	5	3

Solution: Since the distribution is not regular, method of grouping will be used for determination of mode.

Grouping Table

Weight (in Kgs.)	Frequency (1)	(2)	(3)	(4)	(5)	(6)
50	2	[6	[9	[11	[15	
51	4	[11	[14	[19	[19	
52	5	[13	[9	[17	[16	
53	6	[11	[18	[22	[23	
54	8	[16	[8			
55	5					
56	4					
57	7					
58	11					
59	5					
60	3					

Analysis Table

Columns	50	51	52	53	54	55	56	57	58	59	60
1									1		
2									1	1	
3								1	1		
4							1	1	1		
5								1	1	1	
6			1	1	1				1	1	1
Total	0	0	1	1	1	0	1	3	6	3	1

Since the value 58 has occurred maximum number of times, therefore, mode of the distribution is 58 kgs.

(b) When data are in the form of a grouped frequency distribution

The following steps are involved in the computation of mode from a grouped frequency distribution.

- Determination of modal class:** It is the class in which mode of the distribution lies. If the distribution is regular, the modal class can be determined by inspection, otherwise, by method of grouping.

- (ii) **Exact location of mode in a modal class (interpolation formula):** The exact location of mode, in a modal class, will depend upon the frequencies of the classes immediately preceding and following it. If these frequencies are equal, the mode would lie at the middle of the modal class interval. However, the position of mode would be to the left or to the right of the middle point depending upon whether the frequency of preceding class is greater or less than the frequency of the class following it. The exact location of mode can be done by the use of interpolation formula, developed below :

Let the modal class be denoted by $L_m - U_m$, where L_m and U_m denote its lower and the upper limits respectively. Further, let f_m be its frequency and h its width. Also let f_1 and f_2 be the respective frequencies of the immediately preceding and following classes.

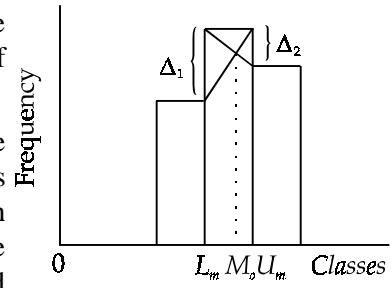


Figure 2.4

We assume that the width of all the class intervals of the distribution are equal. If these are not equal, make them so by regrouping under the assumption that frequencies in a class are uniformly distributed.

Make a histogram of the frequency distribution with height of each rectangle equal to the frequency of the corresponding class. Only three rectangles, out of the complete histogram, that are necessary for the purpose are shown in the above figure.

Let $\Delta_1 = f_m - f_1$ and $\Delta_2 = f_m - f_2$. Then the mode, denoted by M_o , will divide the modal class interval in the ratio $\frac{\Delta_1}{\Delta_2}$. The graphical location of mode is shown in Fig. 2.4.

To derive a formula for mode, the point M_o in the figure, should be such that

$$\begin{aligned} \frac{M_o - L_m}{U_m - M_o} &= \frac{\Delta_1}{\Delta_2} \quad \text{or} \quad M_o \Delta_2 - L_m \Delta_2 = U_m \Delta_1 - M_o \Delta_1 \\ \Rightarrow (\Delta_1 + \Delta_2) M_o &= L_m \Delta_2 + U_m \Delta_1 = L_m \Delta_2 + (L_m + h) \Delta_1 \quad (\text{where } U_m = L_m + h) \\ &= (\Delta_1 + \Delta_2) L_m + \Delta_1 h \end{aligned}$$

Dividing both sides by $\Delta_1 + \Delta_2$, we have

$$M_o = L_m + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h \quad \dots (1)$$

By slight adjustment, the above formula can also be written in terms of the upper limit (U_m) of the modal class.

$$\begin{aligned} M_o &= U_m - h + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h = U_m - \left[1 - \frac{\Delta_1}{\Delta_1 + \Delta_2} \right] \times h \\ &= U_m - \left[\frac{\Delta_2}{\Delta_1 + \Delta_2} \times h \right] \quad \dots (2) \end{aligned}$$

Replacing Δ_1 by $f_m - f_1$ and Δ_2 by $f_m - f_2$, the above equations can be written as

$$M_o = L_m + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h \quad \dots (3)$$

and $M_o = U_m - \frac{f_m - f_2}{2f_m - f_1 - f_2} \times h$ (4)

Note: The above formulae are applicable only to a unimodal frequency distribution.

Example 37: The monthly profits (in Rs) of 100 shops are distributed as follows :

Profit per Shop : 0 - 100	100 - 200	200 - 300	300 - 400	400 - 500	500 - 600
No. of Shops : 12	18	27	20	17	6

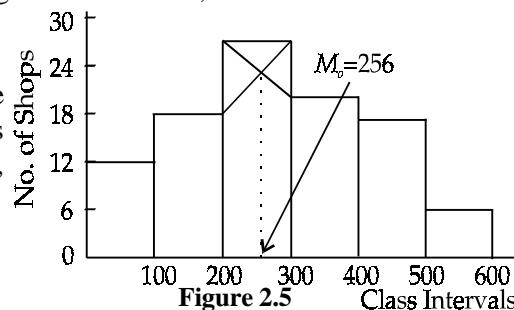
Determine the 'modal value' of the distribution graphically and verify the result by calculation.

Solution: Since the distribution is regular, the modal class would be a class having the highest frequency. The modal class, of the given distribution, is 200 - 300.

Graphical Location of Mode

To locate mode we draw a histogram of the given frequency distribution. The mode is located as shown in Fig. 9.5. From the figure, mode = Rs 256.

Determination of Mode by interpolation formula



Since the modal class is 200 - 300, $L_m = 200$, $\Delta_1 = 27 - 18 = 9$, $\Delta_2 = 27 - 20 = 7$ and $h = 100$.

$$\therefore M_o = 200 + \frac{9}{9+7} \times 100 = \text{Rs } 256.25$$

Example 38: The frequency distribution of marks obtained by 60 students of a class in a college is given below :

Marks : 30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59	60 - 64
Frequency : 3	5	12	18	14	6	2

Find mode of the distribution.

Solution: The given class intervals are first converted into class boundaries, as given in the following table :

Marks : 29.5 - 34.5	34.5 - 39.5	39.5 - 44.5	44.5 - 49.5
Frequency : 3	5	12	18
Marks : 49.5 - 54.5	54.5 - 59.5	59.5 - 64.5	
Frequency : 14	6	2	

We note that the distribution is regular. Thus, the modal class, by inspection, is 44.5 - 49.5.

Further, $L_m = 44.5$, $\Delta_1 = 18 - 12 = 6$, $\Delta_2 = 18 - 14 = 4$ and $h = 5$

$$\therefore \text{Mode} = 44.5 + \frac{6}{6+4} \times 5 = 47.5 \text{ marks}$$

Example 39: Calculate mode of the following data :

Weekly Wages (Rs) : 200-250	250-300	300-350	350-400
No. of Workers : 4	6	20	12
Weekly Wages (Rs) : 400-450	450-500	500-550	550-600
No. of Workers : 33	17	8	2

Solution: Since the frequency distribution is not regular, the modal class will be determined by the method of grouping.

Grouping Table

Weekly wages (in Rs)	f (1)	(2)	(3)	(4)	(5)	(6)
200 - 250	4					
250 - 300	6	10				
300 - 350	20	32	26			
350 - 400	12			45		
400 - 450	33	50	62	65		
450 - 500	17		25		58	
500 - 550	8	10				
550 - 600	2					27

Analysis Table

Columns	300 - 350	350 - 400	400 - 450	450 - 500	500 - 550
1			1		
2			1	1	
3		1	1		
4		1	1	1	
5			1	1	
6	1	1	1		1
<i>Total</i>	1	3	6	3	1

The modal class, from analysis table, is 400 - 500.

Thus, $L_m = 400$, $\Delta_1 = 33 - 12 = 21$, $\Delta_2 = 33 - 17 = 16$ and $h = 50$

$$\text{Hence, mode} = 400 + \frac{21}{37} \times 50 = \text{Rs } 428.38$$

Example 40: Calculate mode of the following distribution :

Weights (lbs.) : below 100 below 110 below 120 below 130 below 140

No. of Students : 4 6 24 46 67

Weights (lbs.) : below 150 below 160 below 170 below 180

No. of Students : 86 96 99 100

Solution: Rewriting the above distribution in the form of a frequency distribution with class limits, we get

Weights (lbs.) : Less than 100 100 - 110 110 - 120 120 - 130 130 - 140

Frequency : 4 2 18 22 21

Weights (lbs.) : 140 - 150 150 - 160 160 - 170 170 - 180

Frequency : 19 10 3 1

We note that there is a concentration of observations in classes 120 - 130 and 130 - 140, therefore, modal class can be determined by the method of grouping.

Grouping Table

Weights (in lbs.)	f (1)	(2)	(3)	(4)	(5)	(6)
less than 100	4					
100 - 110	2	6				
110 - 120	18	40	20			
120 - 130	22	40	43	24		
130 - 140	21	40	43	62	42	
140 - 150	19		29	62	61	
150 - 160	10	13			50	
160 - 170	3		4	14		32
170 - 180	1					

Analysis Table

Measures of Central Tendency

Columns	110 - 120	120 - 130	130 - 140	140 - 150	150 - 160
1		1			
2	1	1	1	1	
3		1	1		
4		1	1	1	
5			1	1	
6	1	1	1		1
<i>Total</i>	2	5	5	3	1

Since the two classes, 120 - 130 and 130 - 140, are repeated maximum number of times in the above table, it is not possible to locate modal class even by the method of grouping.

However, an approximate value of mode is given by the empirical formula:

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean} \quad (\text{See } \S 2.9)$$

Looking at the cumulative frequency column, given in the question, the median class is 130 - 140. Thus, $L_m = 130$, $C = 46$, $f_m = 21$, $h = 10$.

$$\therefore M_d = 130 + \frac{50 - 46}{21} \times 10 = 131.9 \text{ lbs.}$$

Assuming that the width of the first class is equal to the width of second, we can write

Mid-Values (X)	95	105	115	125	135	145	155	165	175	Total
f	4	2	18	22	21	19	10	3	1	100
$u \frac{X - 135}{10}$	4	3	2	1	0	1	2	3	4	
fu	16	6	36	22	0	19	20	9	4	28

$$\text{Thus, } \bar{X} = 135 - \frac{28 \times 10}{100} = 135 - 2.8 = 132.2 \text{ lbs.}$$

Using the values of mean and median, we get

$$M_o = 3 \times 131.9 - 2 \times 132.2 = 131.3 \text{ lbs.}$$

Remarks: Another situation, in which we can use the empirical formula, rather than the interpolation formula, is when there is maximum frequency either in the first or in the last class.

Calculation of Mode when either Δ_1 or Δ_2 is negative

The interpolation formula, for the calculation of mode, is applicable only if both Δ_1 and Δ_2 are positive. If either Δ_1 or Δ_2 is negative, we use an alternative formula that gives only an approximate value of the mode.

We recall that the position of mode, in a modal class, depends upon the frequencies of its preceding and following classes, denoted by f_1 and f_2 respectively. If $f_1 = f_2$, the mode

will be at the middle point which can be obtained by adding $\frac{f_2}{f_1 + f_2} \times h$ to the lower limit

of the modal class or, equivalently, it can be obtained by subtracting $\frac{f_2}{f_1 + f_2} \times h$ from its

upper limit. We may note that $\frac{f_1}{f_1 + f_2} = \frac{f_2}{f_1 + f_2} = \frac{1}{2}$ when $f_1 = f_2$.

Further, if $f_2 > f_1$, the mode will lie to the right of the mid-value of modal class and,

therefore, the ratio $\frac{f_2}{f_1 + f_2}$ will be greater than $\frac{1}{2}$. Similarly, if $f_2 < f_1$, the mode will lie to

the left of the mid-value of modal class and, therefore, the ratio $\frac{f_2}{f_1 + f_2}$ will be less than $\frac{1}{2}$. Thus, we can write an alternative formula for mode as :

$$\text{Mode} = L_m + \frac{f_2}{f_1 + f_2} \times h \text{ or equivalently,} \quad \text{Mode} = U_m - \frac{f_2}{f_1 + f_2} \times h$$

Remarks: The above formula gives only an approximate estimate of mode vis-a-vis the interpolation formula.

Example 41: Calculate mode of the following distribution.

Mid - Values	:	5	15	25	35	45	55	65	75
Frequency	:	7	15	18	30	31	4	3	1

Solution: The mid-values with equal gaps are given, therefore, the corresponding class intervals would be 0 - 10, 10 - 20, 20 - 30, etc.

Since the given frequency distribution is not regular, the modal class will be determined by the method of grouping.

Grouping Table						
Class Intervals	f (1)	(2)	(3)	(4)	(5)	(6)
0 - 10	7					
10 - 20	15	[22]				
20 - 30	18	[48]	[33]	[40]	[63]	
30 - 40	30		[61]			[79]
40 - 50	31	[31]	[35]	[65]		
50 - 60	4				[38]	
60 - 70	3	[4]	[7]			
70 - 80	1					8

Analysis Table

Columns	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
1				1	
2		1	1		
3			1	1	
4			1	1	1
5	1	1	1		
6		1	1	1	
Total	1	3	5	4	1

From the analysis table, the modal class is 30 - 40.

Therefore, $L_m = 30$, $\Delta_1 = 30 - 18 = 12$, $\Delta_2 = 30 - 31 = -1$ (negative) and $h = 10$.

We note that the interpolation formula is not applicable.

$$\text{Mode} = L_m + \frac{f_2}{f_1 + f_2} \times h = 30 + \frac{31}{18+31} \times 10 = 36.33$$

Example 42: The rate of sales tax as a percentage of sales, paid by 400 shopkeepers of a market during an assessment year ranged from 0 to 25%. The sales tax paid by 18% of them was not greater than 5%. The median rate of sales tax was 10% and 75th percentile rate of sales tax was 15%. If only 8% of the shopkeepers paid sales tax at a rate greater than 20% but not greater than 25%, summarise the information in the form of a frequency distribution taking intervals of 5%. Also find the modal rate of sales tax.

Solution: The above information can be written in the form of the following distribution :

<u>Class Intervals (in percentage)</u>	<u>No. of Shopkeepers</u>
0-5	$\frac{18}{100} \times 400 = 72$
5-10	$200 - 72 = 128$
10-15	$300 - 200 = 100$
15-20	$400 - 72 - 128 - 100 = 68$
20-25	$\frac{8}{100} \times 400 = 32$

By inspection, the modal class is 5 - 10.

$$\therefore M_o = 5 + \frac{128 - 72}{128 - 72 + 128 - 100} \times 5 = 8.33\%$$

Example 43: The following table gives the incomplete income distribution of 300 workers of a firm, where the frequencies of the classes 3000 - 4000 and 5000 - 6000 are missing. If the mode of the distribution is Rs 4428.57, find the missing frequencies.

<u>Monthly Income (Rs)</u>	<u>No. of Workers</u>
1000-2000	30
2000-3000	35
3000-4000	?
4000-5000	75
5000-6000	?
6000-7000	30
7000-8000	15

Solution: Let the frequency of the class 3000 - 4000 be f_1 . Then the frequency of the class 5000 - 6000 will be equal to $300 - 30 - 35 - f_1 - 75 - 30 - 15 = 115 - f_1$. It is given that mode = 4428.57, therefore, modal class is 4000 - 5000.

Thus, $L_m = 4000$, $\Delta_1 = 75 - f_1$, $\Delta_2 = 75 - (115 - f_1) = f_1 - 40$ and $h = 1000$.

Using the interpolation formula, we have

$$4428.57 = 4000 + \frac{75 - f_1}{75 - f_1 + f_1 - 40} \times 1000$$

$$\text{or } 428.57 = \frac{75 - f_1}{35} \times 1000 \text{ or } 14.999 = 75 - f_1$$

$$\text{or } f_1 = 75 - 15 = 60 \text{ (taking } 14.999 = 15\text{). Also } f_2 = 115 - 60 = 55$$

Merits and Demerits of Mode

Merits

1. It is easy to understand and easy to calculate. In many cases it can be located just by inspection.
2. It can be located in situations where the variable is not measurable but categorisation or ranking of observations is possible.
3. Like mean or median, it is not affected by extreme observations. It can be calculated even if these extreme observations are not known.
4. It can be determined even if the distribution has open end classes.
5. It can be located even when the class intervals are of unequal width provided that the width of modal and that of its preceding and following classes are equal.

6. It is a value around which there is more concentration of observations and hence the best representative of the data.

Demerits

1. It is not based on all the observations.
2. It is not capable of further mathematical treatment.
3. In certain cases mode is not rigidly defined and hence, the important requisite of a good measure of central tendency is not satisfied.
4. It is much affected by the fluctuations of sampling.
5. It is not easy to calculate unless the number of observations is sufficiently large and reveal a marked tendency of concentration around a particular value.
6. It is not suitable when different items of the data are of unequal importance.
7. It is an unstable average because, mode of a distribution, depends upon the choice of width of class intervals.

2.9 RELATION BETWEEN MEAN, MEDIAN AND MODE

The relationship between the above measures of central tendency will be interpreted in terms of a continuous frequency curve.

If the number of observations of a frequency distribution are increased gradually, then accordingly, we need to have more number of classes, for approximately the same range of values of the variable, and simultaneously, the width of the corresponding classes would decrease. Consequently, the histogram of the frequency distribution will get transformed into a smooth frequency curve, as shown in Fig. 2.6.

For a given distribution, the mean is the value of the variable which is the point of balance or centre of gravity of the distribution. The median is the value such that half of the observations are below it and remaining half are above it. In terms of the frequency curve, the total area under the curve is divided into two equal parts by the ordinate at median. Mode of a distribution is a value around which there is maximum concentration of observations and is given by the point at which peak of the curve occurs.

For a symmetrical distribution, all the three measures of central tendency are equal i.e. $\bar{X} = M_d = M_o$, as shown in Fig. 2.7.

Imagine a situation in which the symmetrical distribution is made asymmetrical or positively (or negatively) skewed by adding some observations of very high (or very low) magnitudes, so that the right hand (or the left hand) tail of the frequency curve gets elongated. Consequently, the three measures will depart from each other. Since mean takes into account the magnitudes of observations, it would be highly affected. Further, since the total number of observations will also increase, the median would also be affected but to a lesser extent than mean. Finally, there would be no change in the position of mode. More specifically, we shall have $M_o < M_d < \bar{X}$, when skewness is positive and $\bar{X} < M_d < M_o$, when skewness is negative, as shown in Fig 2.8.

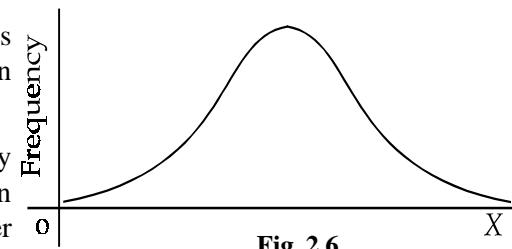


Fig. 2.6

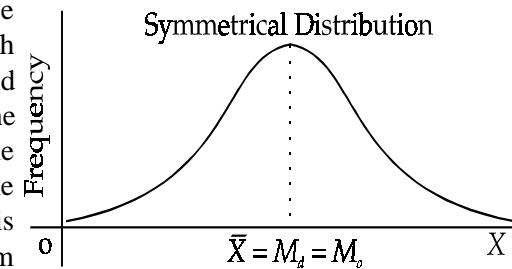


Fig. 2.7

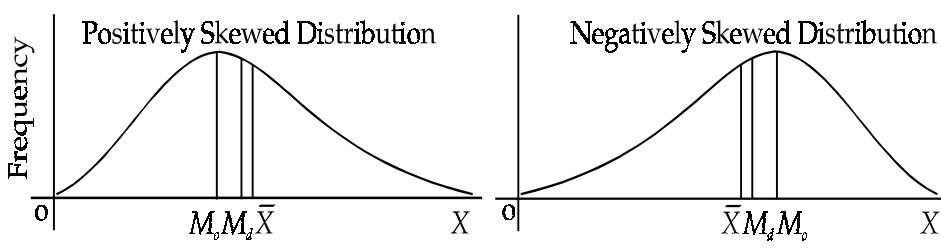


Fig. 2.8

Empirical Relation between Mean, Median and Mode

Empirically, it has been observed that for a moderately skewed distribution, the difference between mean and mode is approximately three times the difference between mean and median, i.e., $\bar{X} - M_o = 3(M_d - \bar{X})$.

This relation can be used to estimate the value of one of the measures when the values of the other two are known.

Example 44:

- The mean and median of a moderately skewed distribution are 42.2 and 41.9 respectively. Find mode of the distribution.
- For a moderately skewed distribution, the median price of men's shoes is Rs 380 and modal price is Rs 350. Calculate mean price of shoes.

Solution:

- Here, mode will be determined by the use of empirical formula.

$$\bar{X} - M_o = 3(M_d - \bar{X}) \quad \text{or} \quad M_o = 3M_d - 2\bar{X}$$

It is given that $\bar{X} = 42.2$ and $M_d = 41.9$

$$\therefore M_o = 3 \times 41.9 - 2 \times 42.2 = 125.7 - 84.4 = 41.3$$

- Using the empirical relation, we can write $\bar{X} = \frac{3M_d - M_o}{2}$

It is given that $M_d = \text{Rs } 380$ and $M_o = \text{Rs. } 350$

$$\therefore \bar{X} = \frac{3 \times 380 - 350}{2} = \text{Rs } 395$$

Example 45: Find mode of the following distribution :

Class Intervals	: 0 - 10 10 - 20 20 - 30 30 - 40 40 - 50
Frequency	: 45 20 14 7 3

Solution: Since the highest frequency occurs in the first class interval, the interpolation formula is not applicable. Thus, mode will be calculated by the use of empirical formula.

Calculation of Mean and Median

Class Intervals	Frequency	c. f.	Mid-Values	u	$\frac{X-25}{10}$	fu
0 - 10	45	45	5	2		90
10 - 20	20	65	15	1		20
20 - 30	14	79	25	0		0
30 - 40	7	86	35	1		7
40 - 50	3	89	45	2		6
Total	89					97

Since $\frac{N}{2} = \frac{89}{2} = 44.5$, the median class is 0 - 10.

$$\therefore M_d = 0 + \frac{44.5 - 0}{45} \times 10 = 9.89$$

$$\text{Also } \bar{X} = 25 - \frac{97 \times 10}{89} = 14.10$$

$$\text{Thus, } M_o = 3M_d - 2\bar{X} = 3 \times 9.89 - 2 \times 14.10 = 1.47$$

Example 46: Estimate mode of the following distribution :

Weekly Wages of Workers (Rs)	:	105-115	115-125	125-135	135-145	145-155
No. of Workers	:	8	15	25	40	62

Solution: We shall use empirical formula for the calculation of mode.

Calculation of \bar{X} and M_d						
Class Intervals	Frequency	c. f.	Mid- Values	u	$\frac{\bar{X} - 130}{10}$	fu
105 - 115	8	8	110	2	16	
115 - 125	15	23	120	1	15	
125 - 135	25	48	130	0	0	
135 - 145	40	88	140	1	40	
145 - 155	62	150	150	2	124	
Total	150					133

Since $\frac{N}{2} = \frac{150}{2} = 75$, the median class is 135 - 145

$$\therefore M_d = 135 + \frac{75 - 48}{40} \times 10 = 135 + 6.75 = 141.75$$

$$\text{Also } \bar{X} = 130 + \frac{133 \times 10}{150} = 138.87$$

$$\text{Thus, } M_o = 3 \times 141.75 - 2 \times 138.87 = 147.51$$

Choice of a Suitable Average

The choice of a suitable average, for a given set of data, depends upon a number of considerations which can be classified into the following broad categories:

(a) Considerations based on the suitability of the data for an average.

(b) Considerations based on the purpose of investigation.

(c) Considerations based on various merits of an average.

(a) **Considerations based on the suitability of the data for an average:**

1. The nature of the given data may itself indicate the type of average that could be selected. For example, the calculation of mean or median is not possible if the characteristic is neither measurable nor can be arranged in certain order of its intensity. However, it is possible to calculate mode in such cases. Suppose that the distribution of votes polled by five candidates of a particular constituency are given as below :

Name of the Candidates	:	A	B	C	D	E
No. of votes polled	:	10,000	5,000	15,000	50,000	17,000

Since the above characteristic, i.e., name of the candidate, is neither measurable nor can be arranged in the order of its intensity, it is not possible to calculate the mean and median. However, the mode of the distribution is D and hence, it can be taken as the representative of the above distribution.

2. If the characteristic is not measurable but various items of the distribution can be arranged in order of intensity of the characteristics, it is possible to locate median in addition to mode. For example, students of a class can be classified into four categories as poor, intelligent, very intelligent and most intelligent. Here the characteristic, intelligence, is not measurable. However, the data can be arranged in ascending or descending order of intelligence. It is not possible to calculate mean in this case.
3. If the characteristic is measurable but class intervals are open at one or both ends of the distribution, it is possible to calculate median and mode but not a satisfactory value of mean. However, an approximate value of mean can also be computed by making certain assumptions about the width of class(es) having open ends.
4. If the distribution is skewed, the median may represent the data more appropriately than mean and mode.
5. If various class intervals are of unequal width, mean and median can be satisfactorily calculated. However, an approximate value of mode can be calculated by making class intervals of equal width under the assumption that observations in a class are uniformly distributed. The accuracy of the computed mode will depend upon the validity of this assumption.

(b) ***Considerations based on the purpose of investigation:***

1. The choice of an appropriate measure of central tendency also depends upon the purpose of investigation. If the collected data are the figures of income of the people of a particular region and our purpose is to estimate the average income of the people of that region, computation of mean will be most appropriate. On the other hand, if it is desired to study the pattern of income distribution, the computation of median, quartiles or percentiles, etc., might be more appropriate. For example, the median will give a figure such that 50% of the people have income less than or equal to it. Similarly, by calculating quartiles or percentiles, it is possible to know the percentage of people having at least a given level of income or the percentage of people having income between any two limits, etc.
2. If the purpose of investigation is to determine the most common or modal size of the distribution, mode is to be computed, e.g., modal family size, modal size of garments, modal size of shoes, etc. The computation of mean and median will provide no useful interpretation of the above situations.

(c) ***Considerations based on various merits of an average:*** The presence or absence of various characteristics of an average may also affect its selection in a given situation.

1. If the requirement is that an average should be rigidly defined, mean or median can be chosen in preference to mode because mode is not rigidly defined in all the situations.
2. An average should be easy to understand and easy to interpret. This characteristic is satisfied by all the three averages.
3. It should be easy to compute. We know that all the three averages are easy to compute. It is to be noted here that, for the location of median, the data must be arranged in order of magnitude. Similarly, for the location of mode, the data should be converted into a frequency distribution. This type of exercise is not necessary for the computation of mean.
4. It should be based on all the observations. This characteristic is met only by mean and not by median or mode.

5. It should be least affected by the fluctuations of sampling. If a number of independent random samples of same size are taken from a population, the variations among means of these samples are less than the variations among their medians or modes. These variations are often termed as sampling variations. Therefore, preference should be given to mean when the requirement of least sampling variations is to be fulfilled. It should be noted here that if the population is highly skewed, the sampling variations in mean may be larger than the sampling variations in median.
6. It should not be unduly affected by the extreme observations. The mode is most suitable average from this point of view. Median is only slightly affected while mean is very much affected by the presence of extreme observations.
7. It should be capable of further mathematical treatment. This characteristic is satisfied only by mean and, consequently, most of the statistical theories use mean as a measure of central tendency.
8. It should not be affected by the method of grouping of observations. Very often the data are summarised by grouping observations into class intervals. The chosen average should not be much affected by the changes in size of class intervals. It can be shown that if the same data are grouped in various ways by taking class intervals of different size, the effect of grouping on mean and median will be very small particularly when the number of observations is very large. Mode is very sensitive to the method of grouping.
9. It should represent the central tendency of the data. The main purpose of computing an average is to represent the central tendency of the given distribution and, therefore, it is desirable that it should fall in the middle of distribution. Both mean and median satisfy this requirement but in certain cases mode may be at (or near) either end of the distribution.

Exercise with Hints

1. The following is the distribution of monthly expenditure on food incurred by a sample of 100 families in a town. Find the modal size of expenditure.

<i>Expenditure (Rs)</i>	:	500-999	1000-1499	1500-1999
<i>No. of families</i>	:	6	25	31
<i>Expenditure (Rs)</i>	:	2000-2499	2500-2999	3000-3499
<i>No. of families</i>	:	26	8	4

Hint: Convert the class intervals into class boundaries.

2. Calculate mode of the following distribution of weekly income of workers of a factory :

<i>Weekly Income</i>	:	0 - 75	75 - 100	100 - 150	150 - 175	175 - 300	300 - 500
<i>No. of Workers</i>	:	9	44	192	116	435	304

Hint: Make class intervals of equal width on the assumption that observations in a class are uniformly distributed. On this basis, the class of 0 - 75 can be written as 0 - 25, 25 - 50 and 50 - 75 each with frequency 3. The class 100 - 150 will be split as 100 - 125 and 125 - 150 each with frequency 96, etc.

3. Calculate the modal marks from the following distribution of marks of 100 students of a class :

<i>Marks (More than)</i>	90	80	70	60	50	40	30	20	10
<i>No. of Students</i>	0	4	15	33	53	76	92	98	100

Hint: Convert 'more than' type frequencies into ordinary frequencies.

4. The following table gives the number of geysers of different sizes (in litres) sold by a company during winter season of last year. Compute a suitable average of the distribution:

<i>Capacity</i>	: less than 5	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30	<i>above 30</i>
<i>Frequency</i>	: 1500	3000	2325	1750	1400	1225	800

Hint: Mode is the most suitable average.

5. Locate a suitable measure of tendency for the following distribution :

<i>Colour of the hair</i>	:	Brown	Black	Grey
<i>No. of Persons</i>	:	200	250	150

Hint: Since the characteristic is neither measurable nor can be arranged in order of magnitude, mode is most suitable.

6. The following table gives the classification of students of a class into various categories according to their level of intelligence. Compute a suitable measures of central tendency.

<i>Characteristics</i>	:	Poor	Intelligent	Very Intelligent	Most Intelligent
<i>No. of Students</i>	:	8	21	25	6

Hint: Median as well as mode.

7. The following table gives the distribution of 200 families according to the number of children :

<i>No. of Children</i>	:	0	1	2	3	4	5	6	7
<i>No. of families</i>	:	12	18	49	62	36	13	7	3

Find \bar{X} , M_d and M_o and interpret these averages.

Hint: \bar{X} represents mean number of children per family. Similarly interpret M_d and M_o .

8. Given below is the income distribution of 500 families of a certain locality :

<i>Monthly Income</i>	:	500 - 1000	1000 - 1500	1500 - 2000	2000 - 2500	2500 - 3000
<i>No. of Families</i>	:	50	210	150	60	30

Find the most suitable average if

- (i) it is desired to estimate average income per family,
- (ii) it is to be representative of the distribution,
- (iii) it is desired to study the pattern of the distribution.

Hint: (i) \bar{X} , (ii) M_o , (iii) M_d , quartiles, percentiles, etc.

9. A distribution of wages paid to foremen would show that, although a few reach very high levels, most foremen are at lower levels of the distribution. The same applies, of course, to most income distributions. If you were an employer, resisting a foreman's claim for an increase of wages, which average would suit your case? Give reasons for supporting your argument. Do you think your argument will be different in case you are a trade union leader?

Hint: An employer should use arithmetic mean because this is the highest average when distribution is positively skewed. Mode will be used by a trade union leader.

10. Atul gets a pocket money allowance of Rs 12 per month. Thinking that this was rather less, he asked his friends about their allowances and obtained the following data which includes his allowance (in Rs) also.

12, 18, 10, 5, 25, 20, 20, 22, 15, 10, 10, 15, 13, 20, 18, 10, 15, 10, 18, 15, 12, 15, 10, 15, 10, 12, 18, 20, 5, 8.

He presented this data to his father and asked for an increase in his allowance as he was getting less than the average amount. His father, a statistician, countered pointing out that Atul's allowance was actually more than the average amount. Reconcile these statements.

Hint: Atul's demand for more pocket money is based on the calculation of arithmetic mean while his father countered his argument on the basis of mode.

2.10 GEOMETRIC MEAN

The geometric mean of a series of n positive observations is defined as the nth root of their product.

Calculation of Geometric Mean

(a) Individual series

If there are n observations, X_1, X_2, \dots, X_n , such that $X_i > 0$ for each i, their geometric mean (GM) is defined as

$$GM = (X_1 \cdot X_2 \cdots \cdot X_n)^{\frac{1}{n}} = \left(\prod_{i=1}^n X_i \right)^{\frac{1}{n}}, \text{ where the symbol P is used to denote the product of observations.}$$

To evaluate GM, we have to use logarithms. Taking log of both sides we have

$$\begin{aligned} \log(GM) &= \frac{1}{n} \log(X_1 \cdot X_2 \cdots \cdot X_n) \\ &= \frac{1}{n} [\log X_1 + \log X_2 + \cdots + \log X_n] = \frac{\sum \log X_i}{n} \end{aligned}$$

Taking antilog of both sides, we have

$$GM = \text{antilog} \left[\frac{\sum \log X_i}{n} \right]$$

This result shows that the GM of a set of observations is the antilog of the arithmetic mean of their logarithms.

Example 47: Calculate geometric mean of the following data :

1, 7, 29, 92, 115 and 375

Solution:

Calculation of Geometric Mean

X	1	7	29	92	115	375	log X
log X	0.0000	0.8451	1.4624	1.9638	2.0607	2.5740	8.9060

$$GM = \text{antilog} \left[\frac{\sum \log X}{n} \right] = \text{antilog} \left[\frac{8.9060}{6} \right] = 30.50$$

(b) Ungrouped Frequency Distribution

If the data consists of observations X_1, X_2, \dots, X_n with respective frequencies f_1, f_2, \dots, f_n , where $\sum_{i=1}^n f_i = N$, the geometric mean is given by=

$$GM = \left[\underbrace{X_1 \cdot X_1 \cdots \cdot X_1}_{f_1 \text{ times}} \underbrace{X_2 \cdots \cdot X_2}_{f_2 \text{ times}} \cdots \underbrace{X_n \cdots \cdot X_n}_{f_n \text{ times}} \right]^{\frac{1}{N}} = \left[X_1^{f_1} \cdot X_2^{f_2} \cdots \cdot X_n^{f_n} \right]^{\frac{1}{N}}$$

Taking log of both sides, we have

Measures of Central Tendency

$$\log(GM) = \frac{1}{N} [\log X_1^{f_1} + \log X_2^{f_2} + \dots + \log X_n^{f_n}]$$

$$= \frac{1}{N} [f_1 \log X_1 + f_2 \log X_2 + \dots + f_n \log X_n] = \frac{\sum_{i=1}^n f_i \log X_i}{N}$$

or $GM = \text{antilog} \left(\frac{1}{N} \sum_{i=1}^n f_i \log X_i \right)$, which is again equal to the antilog of the arithmetic mean of the logarithm of observations.

Example 48: Calculate geometric mean of the following distribution :

X	:	5	10	15	20	25	30
f	:	13	18	50	40	10	6

Solution:

Calculation of GM

X	f	$\log X$	$f \log X$
5	13	0.6990	9.0870
10	18	1.0000	18.0000
15	50	1.1761	58.8050
20	40	1.3010	52.0400
25	10	1.3979	13.9790
30	6	1.4771	8.8626
Total	137		160.7736

$$\therefore GM = \text{antilog} \left[\frac{160.7736}{137} \right] = \text{antilog } 1.1735 = 14.91$$

(c) Continuous Frequency Distribution

In case of a continuous frequency distribution, the class intervals are given. Let X_1, X_2, \dots, X_n denote the mid-values of the first, second nth class interval respectively with corresponding frequencies f_1, f_2, \dots, f_n , such that $\sum f_i = N$. The formula for calculation of GM is same as the formula used for an ungrouped frequency distribution

$$\text{i.e., } GM = \text{antilog} \left[\frac{\sum f_i \log X_i}{N} \right]$$

Example 49: Calculate geometric mean of the following distribution :

Class Intervals	:	5 - 15	15 - 25	25 - 35	35 - 45	45 - 55
Frequencies	:	10	22	25	20	8

Solution:

Calculation of GM

Class	f	Mid-Value (X)	$\log X$	$f \log X$
5 - 15	10	10	1.0000	10.0000
15 - 25	22	20	1.3010	28.6227
25 - 35	25	30	1.4771	36.9280
35 - 45	20	40	1.6020	32.0412
45 - 55	8	50	1.6990	13.5918
Total	85			121.1837

$$GM = \text{antilog} \frac{121.1837}{85} = \text{antilog } 1.4257 = 26.65$$

Weighted Geometric Mean

If various observations, X_1, X_2, \dots, X_n , are not of equal importance in the data, weighted geometric mean is calculated. Weighted GM of the observations X_1, X_2, \dots, X_n with respective weights as w_1, w_2, \dots, w_n is given by :

$$GM = \text{antilog} \left[\frac{\sum w_i \log X_i}{\sum w_i} \right], \text{ i.e., weighted geometric mean of}$$

observations is equal to the antilog of weighted arithmetic mean of their logarithms.

Example 50: Calculate weighted geometric mean of the following data :

Variable (X) :	5	8	44	160	500
Weights (w) :	10	9	3	2	1

How does it differ from simple geometric mean?

Solution:

Calculation of weighted and simple GM

X	Weights (w)	$\log X$	$w \log X$
5	10	0.6990	6.9900
8	9	0.9031	8.1278
44	3	1.6435	4.9304
160	2	2.2041	4.4082
500	1	2.6990	2.6990
Total	25	8.1487	27.1554

$$\text{Weighted GM} = \text{antilog} \frac{27.1554}{25} = \text{antilog } 1.0862 = 12.20$$

$$\text{Simple GM} = \text{antilog} \frac{8.1487}{5} (n=5) = \text{antilog } 1.6297 = 42.63$$

Note that the simple GM is greater than the weighted GM because the given system of weights assigns more importance to values having smaller magnitude.

Geometric Mean of the Combined Group

If G_1, G_2, \dots, G_k are the geometric means of k groups having n_1, n_2, \dots, n_k observations respectively, the geometric mean G of the combined group consisting of $n_1 + n_2 + \dots + n_k$ observations is given by

$$G = \text{antilog} \left[\frac{n_1 \log G_1 + n_2 \log G_2 + \dots + n_k \log G_k}{n_1 + n_2 + \dots + n_k} \right] \text{antilog} \left[\frac{\sum n_i \log G_i}{\sum n_i} \right]$$

Example 51: If the geometric means of two groups consisting of 10 and 25 observations are 90.4 and 125.5 respectively, find the geometric mean of all the 35 observations combined into a single group.

Solution:

$$\text{Combined GM} = \text{antilog} \left[\frac{n_1 \log G_1 + n_2 \log G_2}{n_1 + n_2} \right]$$

Here $n_1 = 10, G_1 = 90.4$ and $n_2 = 25, G_2 = 125.5$

$$\therefore GM = \text{antilog} \left[\frac{10 \log 90.4 + 25 \log 125.5}{35} \right]$$

$$= \text{antilog} \left[\frac{10 \times 1.9562 + 25 \times 2.0986}{35} \right] = \text{antilog } 2.0579 = 114.27$$

To determine the average rate of change of price for the entire period when the rate of change of prices for different periods are given

Let P_0 be the price of a commodity in the beginning of the first year. If it increases by $k_1\%$ in the first year, the price at the end of 1st year (or beginning of second year) is given by

$$P_1 = P_0 + P_0 \frac{k_1}{100} = P_0 \left(1 + \frac{k_1}{100}\right) = P_0(1 + r_1), \text{ where } r_1 = \frac{k_1}{100} \text{ denotes the rate of}$$

increase per rupee in first year. Similarly, if the price changes by $k_2\%$ in second year, the price at the end of second year is given by

$$P_2 = P_1 + P_1 \frac{k_2}{100} = P_1 \left(1 + \frac{k_2}{100}\right) = P_1(1 + r_2)$$

Replacing the value of P_1 as $P_0(1 + r_1)$ we can write

$$P_2 = P_0(1 + r_1)(1 + r_2)$$

Proceeding in this way, if $100r_n\%$ is the rate of change of price in the i th year, the price at the end of n th period, P_n , is given by

$$P_n = P_0(1 + r_1)(1 + r_2) \dots (1 + r_n) \dots (1) \quad \dots (1)$$

Further, let $100r\%$ per year be the average rate of increase of price that gives the price P_n at the end of n years. Therefore, we can write

$$P_n = P_0(1 + r)(1 + r) \dots (1 + r) = P_0(1 + r)_n \dots (2)$$

Equating (1) and (2), we can write

$$(1 + r_n) = (1 + r_1)(1 + r_2) \dots (1 + r_n)$$

$$\text{or } (1 + r) = [(1 + r_1)(1 + r_2) \dots (1 + r_n)]^{\frac{1}{n}} \dots (3)$$

This shows that $(1 + r)$ is geometric mean of $(1 + r_1)$, $(1 + r_2)$, and $(1 + r_n)$.

From (3), we get

$$r = [(1 + r_1)(1 + r_2) \dots (1 + r_n)]^{\frac{1}{n}} - 1 \dots (4)$$

Note: Here r denotes the per unit rate of change. This rate is termed as the rate of increase or the rate of growth if positive and the rate of decrease or the rate of decay if negative.

Example 52: The price of a commodity went up by 5%, 8% and 77% respectively in the last three years. The annual average rise of price is 26% and not 30%. Comment.

Solution: The correct average in this case is given by equation (4), given above.

Let r_1 , r_2 and r_3 be the increase in price per rupee in the respective years.

$$\therefore r_1 = \frac{5}{100} = 0.05, r_2 = \frac{8}{100} = 0.08 \text{ and } r_3 = \frac{77}{100} = 0.77$$

The average rate of rise of price, denoted by r , is given by

$$\begin{aligned} r &= [(1 + r_1)(1 + r_2)(1 + r_3)]^{\frac{1}{3}} - 1 \\ &= [(1 + 0.05)(1 + 0.08)(1 + 0.77)]^{\frac{1}{3}} - 1 = (1.05 \times 1.08 \times 1.77)^{\frac{1}{3}} - 1 \end{aligned}$$

$$\text{Now } \log(1.05 \times 1.08 \times 1.77)^{\frac{1}{3}} = \frac{1}{3}(\log 1.05 + \log 1.08 + \log 1.77) \\ = \frac{1}{3} [0.0212 + 0.0334 + 0.2480] = \frac{1}{3} \times 0.3026 = 0.1009$$

$$\therefore (1.05 \times 1.08 \times 1.77)^{\frac{1}{3}} = \text{antilog } 0.1009 = 1.26$$

$$\text{Thus, } r = 1.26 - 1 = 0.26$$

Also, the percentage rise of price is $100r\% = 26\%$.

Note: 30% is the arithmetic mean of 5%, 8% and 77%, which is not a correct average. This can be verified as below :

If we take the average rise of price as 30% per year, then the price at the end of first year, taking it to be 100 in the beginning of the year, becomes 130.

$$\text{Price at the end of 2nd year} = \frac{130 \times 130}{100} = 169$$

$$\text{Price at the end of 3rd year} = \frac{169 \times 130}{100} = 219.7$$

Similarly, taking the average as 26%, the price at the end of 3rd year

$$= 100 \times \frac{126}{100} \times \frac{126}{100} \times \frac{126}{100} = 200.04$$

Also, the actual price at the end of 3rd year

$$= 100 \times \frac{105}{100} \times \frac{108}{100} \times \frac{177}{100} = 2007. \text{ This price is correctly given by the geometric average and hence, it is the most suitable average in this case.}$$

Average Rate of Growth of Population

The average rate of growth of price, denoted by r in the above section, can also be interpreted as the average rate of growth of population. If P_0 denotes the population in the beginning of the period and P_n the population after n years, using Equation (2), we can write the expression for the average rate of change of population per annum as

$$r = \left(\frac{P_n}{P_0} \right)^{\frac{1}{n}} - 1.$$

Similarly, Equation (4), given above, can be used to find the average rate of growth of population when its rates of growth in various years are given.

Remarks: The formulae of price and population changes, considered above, can also be extended to various other situations like growth of money, capital, output, etc.

Example 53: The population of a country increased from 2,00,000 to 2,40,000 within a period of 10 years. Find the average rate of growth of population per year.

Solution: Let r be the average rate of growth of population per year for the period of 10 years. Let P_0 be initial and P_{10} be the final population for this period.

We are given $P_0 = 2,00,000$ and $P_{10} = 2,40,000$.

$$\therefore r = \left(\frac{P_{10}}{P_0} \right)^{\frac{1}{10}} - 1 = \left(\frac{2,40,000}{2,00,000} \right)^{\frac{1}{10}} - 1$$

$$\text{Now } \left(\frac{24}{20}\right)^{\frac{1}{10}} = \text{antilog} \left[\frac{1}{10}(\log 24 - \log 20) \right]$$

$$= \text{antilog} \left[\frac{1}{10}(1.3802 - 1.3010) \right] = \text{antilog}(0.0079) = 1.018$$

Thus, $r = 1.018 - 1 = 0.018$.

Hence, the percentage rate of growth = $0.018 \times 100 = 1.8\% \text{ p. a.}$

Example 54: The gross national product of a country was Rs 20,000 crores before 5 years. If it is Rs 30,000 crores now, find the annual rate of growth of G.N.P.

Solution: Here $P_5 = 30,000$, $P_0 = 20,000$ and $n = 5$.

$$\therefore r = \left(\frac{30,000}{20,000} \right)^{\frac{1}{5}} - 1$$

$$\text{Now } \left(\frac{3}{2}\right)^{\frac{1}{5}} = \text{antilog} \left[\frac{1}{5}(\log 3 - \log 2) \right] = \text{antilog} \left[\frac{1}{5}(0.4771 - 0.3010) \right]$$

$$= \text{antilog}(0.0352) = 1.084$$

Hence $r = 1.084 - 1 = 0.084$

Thus, the percentage rate of growth of G.N.P. is $8.4\% \text{ p.a}$

Example 55: Find the average rate of increase of population per decade, which increased by 20% in first, 30% in second and 40% in the third decade.

Solution: Let r denote the average rate of growth of population per decade, then

$$r = \left(\frac{120}{100} \times \frac{130}{100} \times \frac{140}{100} \right)^{\frac{1}{3}} - 1 = (1.2 \times 1.3 \times 1.4)^{\frac{1}{3}} - 1$$

$$\text{Now } (1.2 \times 1.3 \times 1.4)^{\frac{1}{3}} = \text{antilog} \left[\frac{1}{3}(\log 1.2 + \log 1.3 + \log 1.4) \right]$$

$$\text{antilog} \left[\frac{1}{3}(0.0792 \ 0.1139 \ 0.1461) \right] = \text{antilog } 0.1131 = 1.297$$

$\therefore r = 1.297 - 1 = 0.297$

Hence, the percentage rate of growth of population per decade is 29.7% .

Suitability of Geometric Mean for Averaging Ratios

It will be shown here that the geometric mean is more appropriate than arithmetic mean while averaging ratios.

Let there be two values of each of the variables x and y , as given below :

x	y	$\text{Ratio} \left(\frac{x}{y} \right)$	$\text{Ratio} \left(\frac{y}{x} \right)$
40	60	$2/3$	$3/2$
20	80	$1/4$	4

Now AM of (x/y) ratios = $\frac{\frac{2}{3} + \frac{1}{4}}{2} = \frac{11}{24}$ and the AM of (y/x) ratios = $\frac{\frac{3}{2} + 4}{2} = \frac{11}{4}$.

We note that their product is not equal to unity.

However, the product of their respective geometric means, i.e., $\frac{1}{\sqrt{6}}$ and $\sqrt{6}$, is equal to unity.

Since it is desirable that a method of average should be independent of the way in which a ratio is expressed, it seems reasonable to regard geometric mean as more appropriate than arithmetic mean while averaging ratios.

Properties of Geometric Mean

- As in case of arithmetic mean, the sum of deviations of logarithms of values from the log GM is equal to zero.

This property implies that the product of the ratios of GM to each observation, that is less than it, is equal to the product the ratios of each observation to GM that is greater than it. For example, if the observations are 5, 25, 125 and 625, their $GM = 55.9$. The above property implies that

$$\frac{55.9}{5} \times \frac{55.9}{25} = \frac{125}{55.9} \times \frac{625}{55.9}$$

- Similar to the arithmetic mean, where the sum of observations remains unaltered if each observation is replaced by their AM, the product of observations remains unaltered if each observation is replaced by their GM.

Merits, Demerits and Uses of Geometric Mean

Merits

- It is a rigidly defined average.
- It is based on all the observations.
- It is capable of mathematical treatment. If any two out of the three values, i.e., (i) product of observations, (ii) GM of observations and (iii) number of observations, are known, the third can be calculated.
- In contrast to AM, it is less affected by extreme observations.
- It gives more weights to smaller observations and vice-versa.

Demerits

- It is not very easy to calculate and hence is not very popular.
- Like AM, it may be a value which does not exist in the set of given observations.
- It cannot be calculated if any observation is zero or negative.

Uses

- It is most suitable for averaging ratios and exponential rates of changes.
- It is used in the construction of index numbers.
- It is often used to study certain social or economic phenomena.

Exercise with Hints

- A sum of money was invested for 4 years. The respective rates of interest per annum were 4%, 5%, 6% and 8%. Determine the average rate of interest p.a.

Hint: $r = \left(\frac{104}{100} \times \frac{105}{100} \times \frac{106}{100} \times \frac{108}{100} \right)^{\frac{1}{4}} - 1$, \therefore average rate of interest = 100r %.

2. The number of bacteria in a certain culture was found to be 4×10^6 at noon of one day. At noon of the next day, the number was 9×10^6 . If the number increased at a constant rate per hour, how many bacteria were there at the intervening midnight?

Hint: The number of bacteria at midnight is GM of 4×10^6 and 9×10^6 .

3. If the price of a commodity doubles in a period of 4 years, what is the average percentage increase per year?

$$\text{Hint: } r = \left(\frac{P_n}{P_0} \right)^{\frac{1}{n}} - 1 = \left(\frac{2}{1} \right)^{\frac{1}{4}} - 1.$$

4. A machine is assumed to depreciate by 40% in value in the first year, by 25% in second year and by 10% p.a. for the next three years, each percentage being calculated on the diminishing value. Find the percentage depreciation p.a. for the entire period.

$$\text{Hint: } 1 - r = \left[(1 - r_1)(1 - r_2)(1 - r_3)^3 \right]^{\frac{1}{5}}.$$

5. A certain store made profits of Rs 5,000, Rs 10,000 and Rs 80,000 in 1965, 1966 and 1967 respectively. Determine the average rate of growth of its profits.

$$\text{Hint: } r = \left(\frac{80,000}{5,000} \right)^{\frac{1}{2}} - 1.$$

6. An economy grows at the rate of 2% in the first year, 2.5% in the second, 3% in the third, 4% in the fourth and 10% in the tenth year. What is the average rate of growth of the economy?

$$\text{Hint: } r = (1.02 \times 1.025 \times 1.03 \times 1.04 \times 1.05 \times 1.06 \times 1.07 \times 1.08 \times 1.09 \times 1.10)^{\frac{1}{10}} - 1.$$

7. The export of a commodity increased by 30% in 1988, decreased by 22% in 1989 and then increased by 45% in the following year. The increase/decrease, in each year, being measured in comparison to its previous year. Calculate the average rate of change of the exports per annum.

$$\text{Hint: } r = (1.30 \times 0.78 \times 1.45)^{\frac{1}{3}} - 1.$$

8. Show that the arithmetic mean of two positive numbers a and b is at least as large as their geometric mean.

Hint: We know that the square of the difference of two numbers is always positive, i.e., $(a - b)^2 \geq 0$. Make adjustments to get the inequality $(a + b)^2 \geq 4ab$ and then get the desired result, i.e., $\text{AM} \geq \text{GM}$.

9. If population has doubled itself in 20 years, is it correct to say that the rate of growth has been 5% per annum?

$$\text{Hint: The annual rate of growth is given by } 100r = 100 \left[\left(2 \right)^{\frac{1}{20}} - 1 \right] = 3.53\%,$$

which is not equal to 5%.

10. The weighted geometric mean of 5 numbers 10, 15, 25, 12 and 20 is 17.15. If the weights of the first four numbers are 2, 3, 5, and 2 respectively, find weight of the fifth number.

$$\text{Hint: Let } x \text{ be the weight of the 5th number, then } \left[10^2 \cdot 15^3 \cdot 25^5 \cdot 12^2 \cdot 20^x \right]^{\frac{1}{12+x}} = 17.15.$$

2.11 HARMONIC MEAN

The harmonic mean of n observations, none of which is zero, is defined as the reciprocal of the arithmetic mean of their reciprocals.

Calculation of Harmonic Mean

(a) Individual series

If there are n observations X_1, X_2, \dots, X_n , their harmonic mean is defined as

$$HM = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \dots + \frac{1}{X_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}}$$

Example 56: Obtain harmonic mean of 15, 18, 23, 25 and 30.

Solution: $HM = \frac{5}{\frac{1}{15} + \frac{1}{18} + \frac{1}{23} + \frac{1}{25} + \frac{1}{30}} = \frac{5}{0.239} = 20.92$ Ans.

(b) Ungrouped Frequency Distribution

For ungrouped data, i.e., each X_1, X_2, \dots, X_n , occur with respective frequency f_1, f_2, \dots, f_n , where $\sum f_i = N$ is total frequency, the arithmetic mean of the reciprocals of observations

is given by $\frac{1}{N} \sum_{i=1}^n \frac{f_i}{X_i}$.

Thus, $HM = \frac{N}{\sum \frac{f_i}{X_i}}$

Example 57: Calculate harmonic mean of the following data :

X	:	10	11	12	13	14
f	:	5	8	10	9	6

Solution:

Calculation of Harmonic Mean

X	10	11	12	13	14	Total
Frequency (f)	5	8	10	9	6	38
$f \frac{1}{X}$	0.5000	0.7273	0.8333	0.6923	0.4286	3.1815

$$\therefore HM = \frac{38}{3.1815} = 11.94$$

(c) Continuous Frequency Distribution

In case of a continuous frequency distribution, the class intervals are given. The mid-values of the first, second n th classes are denoted by X_1, X_2, \dots, X_n . The formula for the harmonic mean is same, as given in (b) above.

Example 58: Find the harmonic mean of the following distribution :

Class Intervals	: 0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
Frequency	: 5	8	11	21	35	30	22	18

Calculation of Harmonic Mean

Class Intervals	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	Total
Frequency (f)	5	8	11	21	35	30	22	18	150
Mid- Values (X)	5	15	25	35	45	55	65	75	
$\frac{f}{X}$	1.0000	0.5333	0.4400	0.6000	0.7778	0.5455	0.3385	0.2400	4.4751

$$\therefore HM = \frac{150}{4.4751} = 33.52 \text{ Ans.}$$

Weighted Harmonic Mean

If X_1, X_2, \dots, X_n are n observations with weights w_1, w_2, \dots, w_n respectively, their weighted harmonic mean is defined as follows :

$$HM = \frac{\sum w_i}{\sum \frac{w_i}{X_i}}$$

Example 59: A train travels 50 kms at a speed of 40 kms/hour, 60 kms at a speed of 50 kms/hour and 40 kms at a speed of 60 kms/hour. Calculate the weighted harmonic mean of the speed of the train taking distances travelled as weights. Verify that this harmonic mean represents an appropriate average of the speed of train.

Solution: $HM = \frac{\sum w_i}{\sum \frac{w_i}{X_i}} = \frac{150}{\frac{50}{40} + \frac{60}{50} + \frac{40}{60}} = \frac{150}{1.25 + 1.20 + 0.67} = 48.13 \text{ kms/hour}$ (1)

Verification : Average speed = $\frac{\text{Total distance travelled}}{\text{Total time taken}}$

We note that the numerator of Equation (1) gives the total distance travelled by train. Further, its denominator represents total time taken by the train in travelling 150 kms,

since $\frac{50}{40}$ is time taken by the train in travelling 50 kms at a speed of 40 kms/hour.

Similarly $\frac{60}{50}$ and $\frac{40}{60}$ are time taken by the train in travelling 60 kms and 40 kms at the speeds of 50 kms./hour and 60 kms/hour respectively. Hence, weighted harmonic mean is most appropriate average in this case.

Example 60: Ram goes from his house to office on a cycle at a speed of 12 kms/hour and returns at a speed of 14 kms/hour. Find his average speed.

Solution: Since the distances of travel at various speeds are equal, the average speed of Ram will be given by the simple harmonic mean of the given speeds.

$$\text{Average speed} = \frac{2}{\frac{1}{12} + \frac{1}{14}} = \frac{2}{0.1547} = 12.92 \text{ kms/hour}$$

Choice between Harmonic Mean and Arithmetic Mean

The harmonic mean, like arithmetic mean, is also used in averaging of rates like price per unit, kms per hour, work done per hour, etc., under certain conditions. To explain the method of choosing an appropriate average, consider the following illustration.

Let the price of a commodity be Rs 3, 4 and 5 per unit in three successive years. If we take A.M. of these prices, i.e., $\frac{3+4+5}{3} = 4$, then it will denote average price when equal quantities of the commodity are purchased in each year. To verify this, let us assume that 10 units of commodity are purchased in each year.

\therefore Total expenditure on the commodity in 3 years = $10 \times 3 + 10 \times 4 + 10 \times 5$.

Also, Average price = $\frac{\text{Total expenditure}}{\text{Total quantity purchased}} = \frac{10 \times 3 + 10 \times 4 + 10 \times 5}{10 + 10 + 10} = \frac{3+4+5}{3}$, which is arithmetic mean of the prices in three years.

Further, if we take harmonic mean of the given prices, i.e., $\frac{3}{\frac{1}{3} + \frac{1}{4} + \frac{1}{5}}$, it will denote the

average price when equal amounts of money are spent on the commodity in three years. To verify this let us assume that Rs 100 is spent in each year on the purchase of the commodity.

$$\therefore \text{Average price} = \frac{\text{Total expenditure}}{\text{Total quantity purchased}} = \frac{300}{\frac{100}{3} + \frac{100}{4} + \frac{100}{5}} = \frac{3}{\frac{1}{3} + \frac{1}{4} + \frac{1}{5}}$$

Next, we consider a situation where different quantities are purchased in the three years. Let us assume that 10, 15 and 20 units of the commodity are purchased at prices of Rs 3, 4 and 5 respectively.

Average price = $\frac{\text{Total expenditure}}{\text{Total quantity purchased}} = \frac{3 \times 10 + 4 \times 15 + 5 \times 20}{10 + 15 + 20}$, which is weighted arithmetic mean of the prices taking respective quantities as weights.

Further, if Rs 150, 200 and 250 are spent on the purchase of the commodity at prices of Rs 3, 4 and 5 respectively, then

Average price = $\frac{150 + 200 + 250}{\frac{150}{3} + \frac{200}{4} + \frac{250}{5}}$, where $\frac{150}{3}$, $\frac{200}{4}$ and $\frac{250}{5}$ are the quantities purchased in respective situations.

The above average price is equal to the weighted harmonic mean of prices taking money spent as weights.

Therefore, to decide about the type of average to be used in a given situation, the first step is to examine the rate to be averaged. It may be noted here that a rate represents a

ratio, e.g., price = $\frac{\text{money}}{\text{quantity}}$, speed = $\frac{\text{distance}}{\text{time}}$, work done per hour = $\frac{\text{work done}}{\text{time taken}}$, etc.

We have seen above that arithmetic mean is appropriate average of prices $\left(\frac{\text{money}}{\text{quantity}} \right)$ when quantities, that appear in the denominator of the rate to be averaged, purchased in different situations are given. Similarly, harmonic mean will be appropriate when sums of money, that appear in the numerator of the rate to be averaged, spent in different situations are given.

To conclude, we can say that the average of a rate, defined by the ratio p/q , is given by the arithmetic mean of its values in different situations if the conditions are given in terms of q and by the harmonic mean if the conditions are given in terms of p . Further, if the conditions are same in different situations, use simple AM or HM and otherwise use weighted AM or HM.

Example 61: An individual purchases three qualities of pencils. The relevant data are given below :

Measures of Central Tendency

Quality	Price per pencil (Rs)	Money Spent (Rs)
A	1.00	50
B	1.50	30
C	2.00	20

Calculate average price per pencil.

Solution: Since different sums of money spent in various situations are given, we shall calculate weighted harmonic mean to calculate average price.

$$\text{Weighted HM} = \frac{\frac{50+30+20}{50+30+20}}{\frac{1.00}{50} + \frac{1.50}{30} + \frac{2.00}{20}} = \frac{100}{50+20+10} = \text{Rs } 1.25$$

Example 62: In a 400 metre athlete competition, a participant covers the distance as given below. Find his average speed.

Speed (Metres per second)

First 80 metres	10
Next 240 metres	7.5
Last 80 metres	10

Solution: Since Speed = $\frac{\text{distance}}{\text{time}}$ and the conditions are given in terms of distance travelled at various speeds, HM will be the appropriate average.

$$\frac{\frac{80+240+80}{80+240+80}}{\frac{10}{80} + \frac{7.5}{240} + \frac{10}{80}} = \frac{400}{8+32+8} = 8.33 \text{ metres/second}$$

Example 63: Peter travelled by a car for four days. He drove 10 hours each day. He drove first day at the rate of 45 kms/hour, second day at the rate of 40 kms/hour, third day at the rate of 38 kms/hour and fourth day at the rate of 37 kms/hour. What was his average speed.

Solution: Since the rate to be averaged is speed = $\left(\frac{\text{distance}}{\text{time}} \right)$ and the conditions are given in terms of time, therefore AM will be appropriate. Further, since Peter travelled for equal number of hours on each of the four days, simple AM will be calculated.

$$\therefore \text{Average speed} = \frac{45+40+38+37}{4} = 40 \text{ kms/hour}$$

Example 64: In a certain factory, a unit of work is completed by A in 4 minutes, by B in 5 minutes, by C in 6 minutes, by D in 10 minutes and by E in 12 minutes. What is their average rate of working? What is the average number of units of work completed per minute? At this rate, how many units of work each of them, on the average, will complete in a six hour day? Also find the total units of work completed.

Solution: Here the rate to be averaged is time taken to complete a unit of work, i.e., $\frac{\text{time}}{\text{units of work done}}$. Since we have to determine the average with reference to a (six hours) day, therefore, HM of the rates will give us appropriate average.

Thus, the average rate of working = $\frac{5}{\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{10} + \frac{1}{12}} = 6.25$ minutes/unit.

The average number of units of work completed per minute = $\frac{1}{6.25} = 0.16$.

The average number of units of work completed by each person = $0.16 \times 360 = 57.6$.

Total units of work completed by all the five persons = $57.6 \times 5 = 288.0$.

Example 65: A scooterist purchased petrol at the rate of Rs 14, 15.50 and 16 per litre during three successive years. Calculate the average price of petrol (i) if he purchased 150, 160 and 170 litres of petrol in the respective years and (ii) if he spent Rs 2,200, 2,500 and 2,600 in the three years.

Solution: The rate to be averaged is expressed as $\frac{\text{money}}{\text{litre}}$

- (i) Since the condition is given in terms of different litres of petrol in three years, therefore, weighted AM will be appropriate.

$$\therefore \text{Average price} = \frac{150 \times 14 + 160 \times 15.5 + 170 \times 16}{150 + 160 + 170} = \text{Rs } 15.21/\text{litre.}$$

- (ii) The weighted HM will be appropriate in this case.

$$\begin{aligned} \text{Average price} &= \frac{2200 + 2500 + 2600}{\frac{2200}{14} + \frac{2500}{15.5} + \frac{2600}{16}} = \frac{7300}{157.14 + 161.29 + 162.50} \\ &= \text{Rs } 15.18/\text{litre} \end{aligned}$$

Merits and Demerits of Harmonic Mean

Merits

1. It is a rigidly defined average.
2. It is based on all the observations.
3. It gives less weight to large items and vice-versa.
4. It is capable of further mathematical treatment.
5. It is suitable in computing average rate under certain conditions.

Demerits

1. It is not easy to compute and is difficult to understand.
2. It may not be an actual item of the given observations.
3. It cannot be calculated if one or more observations are equal to zero.
4. It may not be representative of the data if small observations are given correspondingly small weights.

Relationship among AM, GM and HM

If all the observations of a variable are same, all the three measures of central tendency coincide, i.e., $AM = GM = HM$. Otherwise, we have $AM > GM > HM$.

Example 66: Show that for any two positive numbers a and b, $AM \geq GM \geq HM$.

Solution: The three averages of a and b are :

$$AM = \frac{a+b}{2}, GM = \sqrt{ab} \text{ and } HM = \frac{2}{\frac{1}{a} + \frac{1}{b}} = \frac{2ab}{a+b}.$$

Since the square of the difference between a and b is always a non-negative number, we can write

$$(a - b)^2 \geq 0 \text{ or } a^2 + b^2 - 2ab \geq 0 \text{ or } a^2 + b^2 \geq 2ab.$$

Adding 2ab to both sides, we have

$$a^2 + b^2 + 2ab \geq 4ab \text{ or } (a + b)^2 \geq 4ab$$

$$\text{or } \frac{(a+b)^2}{4} \geq ab \text{ or } \frac{a+b}{2} \geq \sqrt{ab} \quad \dots (1)$$

$$\Rightarrow AM \geq GM \quad \dots (2)$$

Divide both sides of inequality (1) by $\frac{a+b}{2}$, to get $1 \geq \frac{2\sqrt{ab}}{a+b}$

Multiply both sides by \sqrt{ab} , to get $\sqrt{ab} \geq \frac{2ab}{a+b}$

$$\Rightarrow GM \geq HM \quad \dots (3)$$

Combining (2) and (3), we can write

$$AM \geq GM \geq HM$$

Note: The equality sign will hold when $a = b$

Example 67: For any two positive numbers, show that $GM = \sqrt{AM \times HM}$.

Solution: If a and b are two positive numbers, then

$$AM = \frac{a+b}{2}, GM = \sqrt{ab} \text{ and } HM = \frac{2ab}{a+b}$$

$$\text{Now } AM \cdot HM = \frac{a+b}{2} \cdot \frac{2ab}{a+b} = ab = (GM)^2$$

or $GM = \sqrt{AM \times HM}$. Hence the result.

Example 68:

- (a) If AM of two observations is 15 and their GM is 9, find their HM and the two observations.
- (b) Comment on the following :

The AM of 20 observations is 25, $GM = 20$ and $HM = 21$.

Solution:

$$(a) \sqrt{AM \times HM} = GM$$

$$\therefore \sqrt{15 \times HM} = 9 \text{ or } 15 \times HM = 81. \text{ Thus, } HM = 5.4.$$

Let the two observations be X_1 and X_2 . We are given that $\frac{X_1 + X_2}{2} = 15$

$$\text{or } X_1 + X_2 = 30. \quad \dots (1)$$

$$\text{Also } \sqrt{X_1 \cdot X_2} = 9 \text{ or } X_1 \cdot X_2 = 81$$

$$\begin{aligned} \text{We can write } (X_1 - X_2)^2 &= (X_1 + X_2)^2 - 4X_1 X_2 \\ &= 900 - 4 \times 81 = 576 \end{aligned}$$

$$\text{or } X_1 - X_2 = 24 \quad \dots (2)$$

Adding (1) and (2), we get

$$2X_1 = 54, \therefore X_1 = 27. \text{ Also } X_2 = 3$$

- (b) The statement is wrong because HM cannot be greater than GM .

Exercise with Hints

- A train runs 25 miles at a speed of 30 m.p.h., another 50 miles at a speed of 40 m.p.h., then due to repairs of the track, 6 miles at a speed of 10 m.p.h. What should be the speed of the train to cover additional distance of 24 miles so that the average speed of the whole run of 105 miles is 35 m.p.h?

Hint: Let x be the speed to cover a distance of 24 miles,

$$\therefore 35 = \frac{25 + 50 + 6 + 24}{\frac{25}{30} + \frac{50}{40} + \frac{6}{10} + \frac{24}{x}}, \text{ find } x.$$

- Prices per share of a company during first five days of a month were Rs 100, 120, 150, 140 and 50.
 - Find the average daily price per share.
 - Find the average price paid by an investor who purchased Rs 20,000 worth of shares on each day.
 - Find the average price paid by an investor who purchased 100, 110, 120, 130 and 150 shares on respective days.

Hint: Find simple HM in (ii) and weighted AM in (iii).

- Typist A can type a letter in five minutes, B in ten minutes and C in fifteen minutes. What is the average number of letters typed per hour per typist?

Hint: Since we are given conditions in terms of per hour, therefore, simple HM of speed will give the average time taken to type one letter. From this we can obtain the average number of letters typed in one hour by each typist.

$$\text{Simple HM} = \frac{3}{\frac{1}{5} + \frac{1}{10} + \frac{1}{15}} = 8.18 \text{ minutes per letter.}$$

$$\therefore \text{No. of letters typed in 60 minutes} = \frac{60}{8.18} = 7.33$$

- Ram paid Rs 15 for two dozens of bananas in one shop, another Rs 15 for three dozens of bananas in second shop and Rs 15 for four dozens of bananas in third shop. Find the average price per dozen paid by him.

Hint: First find the prices per dozen in three situations and since equal money is spent, HM is the appropriate average.

- A country accumulates Rs 100 crores of capital stock at the rate of Rs 10 crores/year, another Rs 100 crores at the rate of Rs 20 crores/year and Rs 100 crores at the rate of Rs 25 crores/year. What is the average rate of accumulation?

Hint: Since Rs 100 crores, each, is accumulated at the rates of Rs 10, 20 and 25 crores/year, simple HM of these rates would be most appropriate.

- A motor car covered a distance of 50 miles 4 times. The first time at 50 m.p.h., the second at 20 m.p.h., the third at 40 m.p.h. and the fourth at 25 m.p.h. Calculate the average speed.

Hint: Use HM.

- The interest paid on each of the three different sums of money yielding 10%, 12% and 15% simple interest p.a. is the same. What is the average yield percent on the sum invested?

Hint: Use HM.

Quadratic Mean

Measures of Central Tendency

Quadratic mean is the square root of the arithmetic mean of squares of observations.

If X_1, X_2, \dots, X_n are n observations, their quadratic mean is given by

$$QM = \sqrt{\frac{X_1^2 + X_2^2 + \dots + X_n^2}{n}} = \sqrt{\frac{\sum X_i^2}{n}}$$

Similarly, the QM of observations X_1, X_2, \dots, X_n with their respective frequencies as $f_1,$

f_2, \dots, f_n is given by $QM = \sqrt{\frac{\sum f_i X_i^2}{N}}$, where $N = \sum f_i$.

Moving Average

This is a special type of average used to eliminate periodic fluctuations from the time series data.

Progressive Average

A progressive average is a cumulative average which is computed by taking all the available figures in each succeeding years. The average for different periods are obtained as shown below :

$$X_1, \frac{X_1 + X_2}{2}, \frac{X_1 + X_2 + X_3}{3}, \dots, \text{etc.}$$

This average is often used in the early years of a business.

Composite Average

A composite average is an average of various other averages. If for example, $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ are the arithmetic means of k series, their composite average

$$= \frac{\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_k}{k}.$$

Check Your Progress 2.2

- 1 Establish the relation between AM, GM and HM.
2. What is Empirical relation among mean, median and mode.

Notes: (a) Write your answer in the space given below.

- (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
- (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

2.12 LET US SUM UP

Thus we can say that Mean, Median, Mode is the essential phenomena in any statistical analysis. Thus the measure central tendency helps in summarising the data and classify it into simple form.

$$(i) \quad \bar{X} = \frac{\sum f_i X_i}{N} \quad (\text{Simple AM})$$

$$(ii) \quad \bar{X} = A + \frac{\sum f_i d_i}{N} \quad (\text{Short-cut method})$$

$$(iii) \quad \bar{X} = A + h \cdot \frac{\sum f_i u_i}{N} \quad (\text{Step-deviation method})$$

$$(iv) \quad \bar{X}_w = \frac{\sum w_i X_i}{\sum w_i} \quad (\text{Weighted AM})$$

$$(v) \quad \bar{X} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2 + \dots + N_k \bar{X}_k}{N_1 + N_2 + \dots + N_k} \quad (\text{Mean of combined series})$$

$$(vi) \quad M_d = L_m + \frac{\frac{N}{2} - C}{f_m} \times h \quad (\text{Median})$$

$$(vii) \quad Q_i = L_{Q_i} + \frac{\frac{iN}{4} - C}{f_{Q_i}} \quad \text{where } i = 1, 3 \quad (\text{Quartiles})$$

$$(viii) \quad P_k = L_{P_k} + \frac{\frac{kN}{100} - C}{f_{P_k}} \times h \quad (k \text{ th Percentile})$$

$$(ix) \quad GM = \text{Anti log} \left[\frac{\sum f_i \log X_i}{N} \right] \quad (\text{Simple GM})$$

$$(x) \quad GM_w = \text{Anti log} \left[\frac{\sum w_i \log X_i}{\sum w_i} \right] \quad (\text{Weighted GM})$$

$$(xi) \quad G = \text{Anti log} \left[\frac{n_1 \log G_1 + n_2 \log G_2 + \dots + n_k \log G_k}{n_1 + n_2 + \dots + n_k} \right] \quad (\text{GM of the combined series})$$

(xii) $r = [(1+r_1)(1+r_2)(1+r_3)\dots(1+r_n)]^{\frac{1}{n}} - 1$ is average annual rate of growth per unit where r_1, r_2, \dots, r_n are the rates of growth in various years.

$$(xiii) \quad HM = \frac{N}{\sum \frac{f_i}{X_i}} \quad (\text{Simple HM})$$

$$(xiv) \quad HM_w = \frac{\sum w_i}{\sum \frac{w_i}{X_i}} \quad (\text{Weighted HM})$$

2.13 LESSON-END ACTIVITY

The harmonic mean, like arithmetic mean, is also used in averaging of rates like price per unit, kms per hour etc., under certain conditions. Explain the method of choosing an appropriate average between arithmetic mean and harmonic mean.

2.14 KEYWORDS

Mean
Median
Mode
Average
Central Tendency

2.15 QUESTIONS FOR DISCUSSION

1. Write True or False against each of the statement:

- (a) In computation of single arithmetic mean equal importance is given to all the items.
- (b) Median divides the value of variate into two equal parts.
- (c) The value that divides distribution into four equal parts are called Median.
- (d) Mode is that value of the variate which occurs maximum no. of times.
- (e) Harmonic mean is reciprocal of arithmetic mean of their reciprocals.

2. Fill in the blanks :

- (a) is a value which is typical representative of a set of data.
- (b) A measure of is a typical value around which other figures congregate.
- (c) In given observations are arranged in ascending or descending order of magnitude.
- (d) Decile divides distribution into equal parts.
- (e) Mode can be determined in two ways by and by

3. Distinguish between:

- (a) Median and Mode
- (b) Percentile and Decile
- (c) Harmonic Mean and Geometric Mean
- (d) Progressive Average and Composite Average
- (e) Inclusive and Exclusive series

4. Comment on the following:

- (a) Summarisation of data is necessary for any statistical analysis.
- (b) Arithmetic mean is the most popular average in statistics.
- (c) Median is a positional average.
- (d) An average is a substitute for complex group of variables.

2.16 TERMINAL QUESTIONS

1. What is a statistical average? Describe the characteristics of a good statistical average.
2. What are the functions of an average? Discuss the relative merits and demerits of various types of statistical averages.
3. Give the essential requisites of a measure of 'Central Tendency'. Under what circumstances would a geometric mean or a harmonic mean be more appropriate than arithmetic mean?

4. What do you mean by 'Central Tendency'? Describe the advantages and the disadvantages of arithmetic mean and mode.
5. What are the characteristics of an ideal average? How far these are satisfied by the mode and median?
6. Distinguish between a mathematical average and a positional average. Give advantages and disadvantages of each type of average.
7. What do you understand by partition values? Give the definitions of quartiles, deciles and percentiles.
8. "Each average has its own special features and it is difficult to say which one is the best". Explain this statement.
9. Discuss the considerations that determine the selection of a suitable average. Explain by giving one example of each case.
10. Explain the empirical relation between mean, median and mode. What are its uses? Under what circumstances it is expected to hold true?
11. Distinguish between a simple average and a weighted average. Explain with an example the circumstances in which the latter is more appropriate than the former.
12. "An average is a substitute for a complex group of variables but it is not always safe to depend on the substitute alone to the exclusion of individual measurements or groups". Discuss.
13. Show that if all observations of a series are added, subtracted, multiplied or divided by a constant β , the mean is also added, subtracted, multiplied or divided by the same constant.
14. Prove that the algebraic sum of deviations of a given set of observations from their mean is zero.
15. Prove that the sum of squared deviations is least when taken from the mean.
16. The heights of 15 students of a class were noted as shown below. Compute arithmetic mean by using (i) Direct Method and (ii) Short-Cut Method.

S.No. :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ht(cms) :	160	167	174	158	155	171	162	152	156	175	178	167	177	162	153

17. Compute arithmetic mean of the following series :

Marks :	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
No. of Students :	12	18	27	20	17	6

18. Calculate arithmetic mean of the following data :

Mid-Values :	10	12	14	16	18	20
Frequency :	3	7	12	18	10	5

19. Calculate mean from the following data :

Wages (in Rs) :	8-12	14-18	20-24	26-30	32-36	38-42
No. of Workers :	6	10	17	13	3	1

20. Calculate mean marks from the following table :

Marks, less than :	10	20	30	40	50
No. of Students :	25	40	60	75	100

21. The weights (in gms) of 30 articles are given below :

14 16 16 14 22 13 15 24 12 23 14 20 17 21 18 18 19 20 17
 16 15 11 12 21 20 17 18 19 22 23.

Construct a grouped frequency distribution by taking equal class intervals in which the first interval should be 11 - 13 (exclusive). Also find the arithmetic mean.

22. The following information relates to wages of workers in a factory, their total working hours and the average working hours per worker. Calculate the wage per worker and the total wage.

<i>Wages (Rs)</i>	:	50-70	70-90	90-110	110-130	130-150	150-170
<i>Total hours worked</i>	:	72	200	255	154	78	38
<i>Average No. of hours worked per worker</i>	:	9	8	8.5	7	7.8	7.6

23. The monthly salaries of 30 employees of a firm are given below :

69 148 132 118 142 116 139 126 114 100 88 62 77 99 103 144
148 63 104 123 95 80 85 106 123 133 140 134 108 129

The firm gave bonus of Rs 10, 15, 20, 25, 30 and 35 for individuals in the respective salary group; exceeding Rs 60 but not exceeding Rs 75, exceeding Rs 75 but not exceeding Rs 90 and so on up to exceeding Rs 135 but not exceeding Rs 150. Find out the average bonus paid per employee.

24. Find out the missing frequency in the following distribution with mean equal to 30.

<i>Class</i>	:	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
<i>Frequency</i>	:	5	6	10	?	13

25. (a) The following table gives the monthly salary of academic staff of a college. Calculate the simple and weighted arithmetic means of their monthly salary. Which of these averages is most appropriate and why?

	<i>Designation</i>	<i>Monthly Salary</i>	<i>No. of Teachers</i>
(i)	<i>Principal</i>	4500	1
(ii)	<i>Reader</i>	3700	5
(iii)	<i>Senior - Lecturer</i>	3000	15
(iv)	<i>Lecturer</i>	2200	25

- (b) The sum of deviations of a certain number of observations from 12 is 166 and the sum of deviations of these observations from 16 is 54. Find the number of observations and their mean.

26. Twelve persons gambled on a certain night. Seven of them lost at an average rate of Rs 10.50 while remaining five gained at an average of Rs 13.00. Is the information given above is correct? If not, why?

27. The incomes of employees in an industrial concern are given below. The total income of ten employees in the class over Rs 250 is Rs 3,000. Compute mean income. Every employee belonging to the top 25% of the earners is required to pay 1% of his income to workers' relief fund. Estimate the contribution to this fund.

<i>Income (Rs)</i>	:	0-50	50-100	100-150	150-200	200-250	250 and above
<i>Frequency</i>	:	90	150	100	80	70	10

28. Comment on the performance of the students of three universities given below:

<i>Courses of Study</i>	<i>Bombay University</i>		<i>Calcutta University</i>		<i>Madras University</i>	
	<i>Pass%</i>	<i>No. of Students</i>	<i>Pass%</i>	<i>No. of Students</i>	<i>Pass%</i>	<i>No. of Students</i>
<i>M. A.</i>	71	300	82	200	81	200
<i>M. Com.</i>	83	400	76	300	76	350
<i>M. Sc.</i>	66	300	60	700	73	200
<i>B. A.</i>	73	500	73	600	74	450
<i>B. Com.</i>	74	200	76	700	58	200
<i>B. Sc.</i>	65	300	65	300	70	700

29. (a) Compute the weighted arithmetic mean of the indices of various groups as given below:

<i>Group</i>	<i>Index</i>	<i>Weight</i>
<i>Food</i>	120	4
<i>Clothing</i>	130	2
<i>Housing</i>	150	2
<i>Education of Children</i>	100	1
<i>Miscellaneous</i>	160	1

- (b) A cumulative frequency distribution has 65 as the mid-value of its last class interval. The cumulative frequencies of the first, second seventh classes are 5, 21, 45, 72, 85, 94 and 100 respectively. If all the class intervals are of equal width of 10 units, write down the relevant frequency distribution. Also calculate its mean and median.
30. A distribution consists of three components each with total frequency of 200, 250 and 300 and with means of 25, 10 and 15 respectively. Find out the mean of the combined distribution.
31. Find the average number of children per family for the sub-groups separately as well as combined as a whole.

<i>Sub - group I</i>		<i>Sub - group II</i>	
<i>No. of Children</i>	<i>No. of families</i>	<i>No. of Children</i>	<i>No. of families</i>
0	10	4 - 5	20
1	50	6 - 7	12
2	60	8 - 9	4
3	40	10 - 11	4

32. (a) The mean of a certain number of items is 20. If an observation 25 is added to the data, the mean becomes 21. Find the number of items in the original data.
 (b) The mean age of a combined group of men and women is 30 years. If the mean age of the men's group is 32 years and that for the women's group is 27 years, find the percentage of men and women in the combined group.
33. The average age of 40 students entering B.A. (Honours) Economics first year in a college was 19 years. Out of this only 25 students passed the third year examination. If the average age of these 25 students is 22.5 years, find the average age of the remaining students.
34. Fifty students took a test. The result of those who passed the test is given below:
- | | | | | | | | |
|------------------------|---|---|----|---|---|---|---|
| <i>Marks</i> | : | 4 | 5 | 6 | 7 | 8 | 9 |
| <i>No. of Students</i> | : | 8 | 10 | 9 | 6 | 4 | 3 |
- If the average marks for all the 50 students was 5.16, find the average marks of those who failed.
35. A person had 7 children. The average age of the children was 14 years when one of the child died at the age of 8 years. What will be the average age of the remaining children after five years of this death?
36. The mean marks of 100 students was calculated as 40. Later on it was discovered that a score 53 was misread as 83. Find the correct mean.
37. An examination was held to decide the award of a scholarship. The weights given to various subjects were different. Only three applicants for the scholarship obtained over 50% marks in aggregate. The marks were as follows :

<i>Subjects</i>	<i>Weights</i>	<i>% Marks of A</i>	<i>% Marks of B</i>	<i>% Marks of C</i>
<i>Cost Accounting</i>	5	70	65	90
<i>Statistics</i>	4	63	80	75
<i>Business Law</i>	2	50	40	65
<i>Economics</i>	3	55	50	40
<i>Insurance</i>	1	60	40	38

Of the candidates, the one getting the highest average marks is to be awarded the scholarship. Determine, who will get it?

Measures of Central Tendency

38. The number of fully formed tomatoes on 100 plants were counted with the following results :

2	plants had	0	tomatoes
5	"	1	"
7	"	2	"
11	"	3	"
18	"	4	"
24	"	5	"
12	"	6	"
8	"	7	"
6	"	8	"
4	"	9	"
3	"	10	"

- (i) How many tomatoes were there in all?
(ii) What was the average number of tomatoes per plant?
39. (a) The average income of 300 employees of a company is Rs 1,800 p.m. Due to rise in prices the company owner decided to give ad-hoc increase of 25% of the average income to each of the 25% lowest paid employees, 10% of the average income to each of the 10% highest paid employees and 15% to each of the remaining employees. Find out the amount of money required for ad hoc increase and also the average income of an employee after this increase.
(b) The frequency distribution of the number of casual leave taken by the employees of a firm in a particular year is given below in which one entry marked as '?' is missing. Determine the missing value if the average number of casual leave taken by an employee is 8.5.

No. of Casual leave taken	:	0	4	5	?	9	10	12
No. of Employees	:	8	35	40	65	79	91	82

40. The mean salary paid to 1,000 employees of an establishment was found to be Rs 180.40. Later on, after disbursement of salary, it was discovered that the salaries of two employees were wrongly entered as Rs 297 and Rs 165 instead of Rs 197 and 185 respectively. Find the correct mean salary.
41. The following variations were recorded in the measurements of parts by a machine:

Variations from the Standard (mm.)	No. of parts
10 to 15	1
5 to 10	3
0 to 5	20
5 to 0	25
10 to 5	22
15 to 10	17
20 to 15	13
25 to 20	10
30 to 25	7
35 to 30	2

- (i) Find average variations.
(ii) What proportion fell within a range of 5 mm. either way of the standard?

- (iii) If those which fall more than 10 mm. apart from the standard are classified as bad, what percentage of the parts are bad?
 - (iv) Which stretch of 15 mm. contains the greatest number of parts and what fraction of the total fall inside this stretch?
42. (a) The average monthly production of a certain factory for the first ten months of a year was 3,500 units. Due to workers' unrest in the last two months, the average monthly production for the whole year came down to 3,200 units. Find the average monthly production of the last two months.
- (b) The average sales of a balloon seller on the first five days (i.e., Monday to Friday) of a particular week was Rs 50 and his average sales for the entire week was Rs 70. If his sales on Sunday were 40% higher than his sales on Saturday, find his sales on each of the last two days, i.e., on Saturday and Sunday.
43. Determine median from the following data :
- 30, 37, 54, 58, 61, 64, 31, 34, 52, 55, 62, 28, 47, 55, 60
44. Locate median of the following data:
- 65, 85, 55, 75, 96, 76, 65, 60, 40, 85, 80, 125, 115, 40
45. Locate M_d , Q_1 , Q_3 , D_3 , D_6 , P_{20} , P_{40} , P_{85} and P_{90} from the following data :

S. No.	Marks	S. No.	Marks	S. No.	Marks
1	17	7	41	13	11
2	32	8	32	14	15
3	35	9	10	15	35
4	33	10	18	16	23
5	15	11	20	17	38
6	21	12	22	18	12

46. In a class of 16 students, the following are the marks obtained by them in statistics. Find out the lower quartile, upper quartile, seventh decile and thirty-fifth percentile.

S. No. :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Marks :	5	12	17	23	28	31	37	41	42	49	54	58	65	68	17	77

47. Locate M_d , Q_1 , Q_3 , D_4 , D_7 , P_{26} , P_{45} , P_{66} , P_{70} and P_{79} from the following data :

Age of Children (in years) :	6	7	8	9	10	11	12	13	14	15
No. of Children :	32	33	39	43	58	59	52	38	33	13

48. Find median from the series given below :

Marks (less than) :	5	10	15	20	25	30	35	40	45	50
No. of Students :	5	13	28	53	83	105	123	135	142	145

49. Calculate median from the following table :

Wages (more than) :	30	40	50	60	70	80	90
No. of Workers :	58	46	40	31	16	5	0

50. Calculate median from the following data :

Class :	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50
Frequency :	5	8	5	22	20	25	19	25	6	5

51. Compute median from the following data :

Class :	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
Frequency :	15	8	17	29	7	4

52. Find out median from the following :

<i>No. of Workers</i>	:	1 - 5	6 - 10	11 - 15	16 - 20	21 - 25
<i>No. of factories</i>	:	3	8	13	11	5

53. Calculate median income for the following distribution :

<i>Income (Rs)</i>	:	40 - 44	45 - 49	50 - 54	55 - 59	60 - 64	65 - 69	70 - 74
<i>No. of Persons</i>	:	2	7	10	12	8	3	3

54. With the help of the following figures, prepare a cumulative frequency curve and locate the median and quartiles:

<i>Marks Obtained</i>	:	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
<i>No. of Students</i>	:	10	12	20	18	10

55. Draw a cumulative frequency curve from the following data and find out the median and both quartiles:

<i>Class</i>	:	1 - 5	6 - 10	11 - 15	16 - 20	21 - 25	26 - 30	31 - 35	36 - 40	41 - 45
<i>Frequency</i>	:	7	10	16	32	24	18	10	5	1

56. Calculate median and both quartiles from the following data :

<i>Age</i>	:	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59
<i>No. of Persons</i>	:	50	70	100	180	150	120	70	60

57. Calculate the quartiles, D_7 and P_{85} from the following data :

<i>Class</i>	:	Less than 100	100 - 250	250 - 400	400 - 500	500 - 550
<i>Frequency</i>	:	85	100	175	74	66
<i>Class</i>	:	550 - 600	600 - 800	800 - 900	900 - 1000	
<i>Frequency</i>	:	35	5	18	2	

58. Calculate arithmetic mean and median from the data given below :

<i>Income in Rs (less than)</i>	:	80	70	60	50	40	30	20	10
<i>No. of Workers</i>	:	100	90	80	60	32	20	13	5

59. Calculate mean and median from the following series :

<i>Class Intervals</i>	:	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
<i>Frequency</i>	:	15	20	18	27	20

60. Calculate mean and median from the following table :

<i>Price (Rs)</i>	:	10 - 20	10 - 30	10 - 40	10 - 50	10 - 60	10 - 70	10 - 80	10 - 90
<i>Frequency</i>	:	4	16	56	97	124	137	146	150

61. Compute mean and median from the following data :

<i>Marks obtained</i>	:	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69	70 - 79	80 - 89
<i>No. of Students</i>	:	0	5	10	17	18	30	10	8	2

62. Calculate mean and median of the following distribution :

<i>Size</i>	:	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24	24 - 28	28 - 32
<i>Frequency</i>	:	5	7	9	17	15	14	6	0

63. Following is the distribution of marks obtained by 50 students in 'mercantile law'. Calculate median marks. If 60% of the students pass this test, find the minimum marks obtained by a passed candidate.

<i>Marks (more than)</i>	:	0	10	20	30	40	50
<i>No. of Students</i>	:	50	46	40	20	10	3

64. Estimate the number of first, second and third divisioners and the number of failures from the following data. First division is awarded at 60 or more marks, second division at 50 and above but less than 60, third division at 36 or more but less than 50 and those securing less than 36 are failures.

<i>Marks (out of 100)</i>	: 0 - 20	20 - 40	40 - 60	60 - 80	80 and above
<i>No. of Students</i>	: 18	30	65	25	12

65. Following relate to the weekly wages (in Rs) of workers of a factory :
- 100, 75, 79, 80, 110, 93, 109, 84, 95, 77, 100, 89, 84, 81, 106, 96, 94, 83, 95, 78, 101, 99, 83, 89, 102, 97, 93, 82, 97, 80, 102, 96, 87, 99, 107, 99, 97, 80, 98, 93, 106, 94, 88, 104, 103, 100, 98, 84, 100, 96, 86, 93, 89, 100, 101, 106, 92, 86, 105, 97, 82, 92, 75, 103, 101, 103, 100, 88, 106, 98, 87, 90, 76, 104, 101, 107, 97, 91, 103, 98, 109, 86, 76, 107, 88, 107, 88, 93, 85, 98, 104, 78, 79, 110, 94, 108, 86, 95, 84, 87.

Prepare a frequency distribution by taking class intervals as 75 - 80, 80 - 85, etc. and locate its median and the two quartiles.

66. Find an appropriate average for the following distribution :

<i>Weekly Income (in Rs)</i>	<i>No. of families</i>
Below 100	50
100 - 200	500
200 - 300	555
300 - 400	100
400 - 500	3
500 and above	2

67. In the frequency distribution of 100 families given below, the number of families corresponding to weekly expenditure groups 200 - 400 and 600 - 800 are missing. However, the median of the distribution is known to be Rs 500. Find the missing frequencies.

<i>Expenditure</i>	: 0 - 200	200 - 400	400 - 600	600 - 800	800 - 1000
<i>No. of families</i>	: 14	?	27	?	15

68. Find median from the following distribution :

<i>X</i>	:	1	2	3	4	5 - 9	10 - 14	15 - 19	20 - 25
<i>f</i>	:	5	10	16	20	30	15	8	6

69. The following is the monthly wage distribution of a certain factory :

<i>Wages (Rs)</i>	:	50-80	80-100	100-110	110-120	120-130
<i>No. of Workers</i>	:	50	120	200	250	170
<i>Wages (Rs)</i>	:	130-150	150-170	170-200		
<i>No. of Workers</i>	:	130	60	20		

- (a) Find the median wage.
 (b) A fund is to be raised and it is decided that the workers getting less than Rs 120 should contribute 5% of their wages and those getting Rs 120 or more should contribute 10% of their wages. What sum should be collected?

70. Determine the mode of the following data :

58, 60, 31, 62, 48, 37, 78, 43, 65, 48

71. Locate mode of the following series :

<i>S. No.</i>	:	1	2	3	4	5	6	7	8	9	10	11	12
<i>Age</i>	:	9	7	4	9	10	8	4	10	5	8	15	8

72. Determine whether there is any mode in the following series :

Measures of Central Tendency

<i>S. No.</i>	: 1	2	3	4	5	6	7	8
<i>Size</i>	: 10	10	10	10	10	10	10	10

73. The number of calls received in 240 successive one minute intervals at an exchange are shown in the following frequency distribution. Calculate mode:

<i>No. of calls</i>	: 0	1	2	3	4	5	6	7
<i>Frequency</i>	: 14	21	25	43	51	35	39	12

74. Calculate mode from the following data :

<i>Midpoints</i>	: 1	2	3	4	5	6	7	8
<i>Frequency</i>	: 5	50	45	30	20	10	15	5

75. Calculate mode from the following series :

<i>Class Intervals</i>	: 10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69
<i>Frequency</i>	: 4	6	8	5	4	2

76. Calculate mode of the following frequency distribution :

<i>Marks</i>	: 0 - 6	6 - 12	12 - 18	18 - 24	24 - 30	30 - 36
<i>No. of Students</i>	: 12	24	36	38	37	6

77. Calculate mode from the following distribution :

<i>Marks (less than)</i>	: 7	14	21	28	35	42	49
<i>No. of Students</i>	: 20	25	33	41	45	50	52

78. Calculate median and mode from the following data :

<i>Size</i>	: 10 - 20	10 - 30	10 - 40	10 - 50	10 - 60	10 - 70	10 - 80	10 - 90
<i>Frequency</i>	: 4	16	56	97	124	137	146	150

79. Calculate \bar{X} and M_o from the following distribution :

<i>Class Intervals</i>	: 6 - 10	11 - 15	16 - 20	21 - 25	26 - 30
<i>Frequency</i>	: 20	30	50	40	10

80. Find out mode of the following data graphically and check the result by calculation:

<i>Size</i>	: 0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8	8 - 9	9 - 10	10 - 11
<i>Frequency</i>	: 3	7	9	15	25	20	14	12	8	6	2

81. (a) Construct a frequency distribution of the marks obtained by 50 students in economics as given below :

42, 53, 65, 63, 61, 47, 58, 60, 64, 45, 55, 57, 82, 42, 39, 51, 65, 55, 33, 70, 50, 52, 53, 45, 45, 25, 36, 59, 63, 39, 65, 30, 45, 35, 49, 15, 54, 48, 64, 26, 75, 20, 42, 40, 41, 55, 52, 46, 35, 18. (Take the first class interval as 10 - 20).

(b) Calculate mode of the above distribution.

82. The monthly profits (in Rs) of 100 shops are distributed as follows :

<i>Profits</i>	: 0 - 100	100 - 200	200 - 300	300 - 500	500 - 600	600 - 800
<i>No. of Shops</i>	: 19	21	30	40	10	12

Calculate mode of the distribution.

83. The mode of the following incomplete distribution of weights of 160 students is 56. Find the missing frequencies.

<i>Weights (kgs)</i>	: 30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90
<i>No. of Students</i>	: 20	36	?	?	15	5

84. Calculate mean, median and mode from the following table :

<i>Wages (Rs)</i>	<i>No. of Persons</i>
<i>Less than 8</i>	5
<i>Less than 16</i>	12
8-24	29
<i>24 and above</i>	31
32-40	8
<i>40 and above</i>	19
48-56	5

85. (a) In a moderately skewed distribution, the arithmetic mean is 10 and mode is 7. Find median.
- (b) In a moderately asymmetrical distribution, the mean is 25 and the median is 23.5. Find mode.
86. Find geometric mean from the following daily income (in Rs) of 10 families: 85, 70, 15, 75, 500, 8, 45, 250, 40 and 36.
87. Calculate geometric mean of the following distribution :
- | <i>Marks (less than)</i> | : 10 20 30 40 50 |
|--------------------------|-------------------------------|
| <i>No. of Students</i> | : 12 27 72 93 100 |
88. The value of a machine depreciates at a constant rate from the cost price of Rs 1,000 to the scrap value of Rs 100 in ten years. Find the annual rate of depreciation and the value of the machine at the end of one, two, three years.
89. Calculate weighted GM from the following data :

<i>Items</i>	<i>Weights</i>	<i>Price Index</i>
<i>Wheat</i>	10	135
<i>Milk</i>	5	140
<i>Sugar</i>	2	160
<i>Eggs</i>	6	120

90. The price of a commodity increased by 12% in 1986, by 30% in 1987 and by 15% in 1988. Calculate the average increase of price per year.
91. The population of a city was 30 lakh in 1981 which increased to 45 lakh in 1991. Determine the rate of growth of population per annum. If the same growth continues, what will be the population of the city in 1995.
92. The value of a machine depreciated by 30% in 1st year, 13% in 2nd year and by 5% in each of the following three years. Determine the average rate of depreciation for the entire period.
93. The following table gives the diameters of screws obtained in a sample enquiry. Calculate mean diameter by using geometric average.

<i>Diameter (mm)</i>	: 130 135 140 145 146 148 149 150 157
<i>No. of Screws</i>	: 3 4 6 6 3 5 2 1 1

94. (a) The price of a commodity doubles in a period of 5 years. What will be the average rate of increase per annum.
- (b) If a sum of Rs 1,500 is invested at 15% rate of interest compounded annually, determine the amount after 5 years.

95. (a) Find the average rate of increase per decade in the population which increased by 10% in the first decade, by 20% in the second and by 40% in the third.
- (b) The price of a commodity increased by 10% in 1st year, by 15% in 2nd year and decreased by 10% in 3rd year. Determine the average change of price after 3 years.
96. The following table gives the marks obtained by 70 students in mathematics. Calculate arithmetic and geometric means:

<i>Marks (more than)</i>	:	80	70	60	50	40	30	20
<i>No. of Students</i>	:	0	7	18	40	40	63	70

97. The population of a city has grown in the following manner :

<i>Years</i>	:	1951	1961	1971	1981	1991
<i>Population (lacs)</i>	:	10	13	15.5	20.8	30.5

Find the average growth per decade.

98. The geometric means of three groups consisting of 15, 20 and 23 observations are 14.5, 30.2 and 28.8 respectively. Find geometric mean of the combined group.
99. A sum of money was invested for 3 years. The rates of interest in the first, second and third year were 10%, 12% and 14% respectively. Determine the average rate of interest per annum.
100. The weighted geometric mean of four numbers 8, 25, 17 and 30 is 15.3. If the weights of first three numbers are 5, 3 and 4 respectively, find the weight of the fourth number.
101. The annual rates of growth of output of a factory in five years are 5.0, 6.5, 4.5, 8.5 and 7.5 percent respectively. What is the compound rate of growth of output per annum for the period?
102. (a) A man invested Rs 1,000, Rs 12,000 and Rs 15,000 at the respective rates of return of 5%, 14% and 13% p.a. respectively. Determine his average rate of return per annum.
- (b) The arithmetic and the geometric means of two numbers are 20.5 and 20 respectively. Find the numbers.
103. (a) Calculate the harmonic mean of the following data :
- 9, 5, 2, 10, 15, 35, 20, 24, 21
- (b) Calculate HM of the following items :
- 1.0, 1.5, 15.0, 250, 0.5, 0.05, 0.095, 1245, 0.009
104. Calculate \bar{X} , GM and HM and verify that $\bar{X} > \text{GM} > \text{HM}$.
- | | | | | | | |
|------------------------|---|--------|---------|---------|---------|---------|
| <i>Class Intervals</i> | : | 5 - 15 | 15 - 25 | 25 - 35 | 35 - 45 | 45 - 55 |
| <i>Frequency</i> | : | 6 | 9 | 15 | 8 | 4 |
105. Four typists take 15, 10, 8, 7 minutes respectively to type a letter. Determine the average time required to type a letter if
- (a) Four letters are to be typed by each typist.
- (b) Each typist works for two hours.
106. (a) A person spends Rs 60 for oranges costing Rs 10 per dozen and another Rs 70 for oranges costing Rs 14 per dozen. What is the average price per dozen paid by him?
- (b) Three mechanics take 10, 8, and 6 hours respectively to assemble a machine. Determine the average number of hours required to assemble one machine.

107. At harvesting time, a farmer employed 10 men, 20 women and 16 boys to lift potatoes. A woman's work was three quarters as effective as that of a man, while a boy's work was only half. Find the daily wage bill if a man's rate was Rs 24 per day and the rates for the women and boys were in proportion to their effectiveness. Calculate the average daily rate for the 46 workers.
108. Saddam takes a trip which entails travelling 1,350 kms by train at a speed of 60 kms/hr, 630 kms by aeroplane at 350 kms/hr, 4,500 kms by ship at 25 kms/hr and 20 kms by car at 30 kms/hr. What is the average speed for the entire journey?
109. (a) A man travels from Lucknow to Kanpur, a distance of 80 kms, at a speed of 45 kms/hr. From Kanpur he goes to Etawah, a distance of 165 kms, at a speed of 65 kms/hr and from Etawah he comes back to Lucknow, along the same route, at a speed of 60 kms/hr. What is his average speed for the entire journey?
 (b) If refills for 5 rupees are purchased at 40 paise each and for another 5 rupees are purchased at 60 paise each, the average price would be 48 paise and not 50 paise. Explain and verify.
110. (a) An aeroplane travels distances of 2,500, 1,200, and 500 kms at the speeds of 500, 400 and 250 kms/hour respectively. Find the average speed for the entire trip, commenting upon the choice of your average.
 (b) A train goes from Delhi to Agra in four hours at speeds of 25, 60, 80 and 40 kms/hour in each successive hour respectively. Find the average speed of the train and verify your answer.
111. A can do a unit of work in 10 minutes, B in 18 minutes and C in 20 minutes. Find their average rate of working when :
 (i) A works for 8 hours, B for 9 hours and C for 10 hours per day.
 (ii) Each of them have to complete 40 units of work per day.
 Also determine the total units of work done per day in each of the above situations and verify your answer.
112. Choose an appropriate average to find the average price per kg., for the following data:

Articles	Qty Purchased	Rate (in gms./rupee)
Wheat	5 kg.	250
Rice	3 kg.	150
Sugar	1 kg.	100
Pulses	2 kg.	90

113. Calculate the weighted harmonic mean of the following data :

$$\begin{array}{l} X : 3 \quad 10 \quad 25 \quad 40 \\ w : 6 \quad 3 \quad 4 \quad 1 \end{array}$$

Now change the weights as 12, 6, 8 and 2 respectively and recalculate the weighted harmonic mean. What do you conclude?

114. (a) The speeds of various buses of a company plying on the same route was found to be as given below :

$$\begin{array}{l} \text{Speed (in miles/hour)} : 12 \quad 15 \quad 18 \\ \text{No. of Buses} : 3 \quad 5 \quad 2 \end{array}$$

Find the average speed of the 10 buses.

- (b) Find mean daily earnings from the following data :

50 men get at the rate of Rs 50 per man per day

35	"	60	"
25	"	75	"
10	"	100	"

115. A college canteen sells tea for 75 paise per cup, coffee for Rs 1.50 per cup and bread pakora for Rs 2 per plate. If on a particular day, it sold tea worth Rs 150, coffee worth Rs 165 and bread pakora worth Rs 200, what is the average price per item sold?

116. A firm of readymade garments makes both men's and women's shirts. Its profit average 6% of sales ; its profit in men's shirts average 8% of sales. If the share of women's shirts in total sales is 60%, find the average profit as a percentage of the sales of women's shirts.

117. Which of the averages will be most suitable in the following circumstances?

- (i) Average rate of growth of population in a given period.
- (ii) Average number of children in a family.
- (iii) Average size of oranges on a tree.
- (iv) Average speed of work.
- (v) Average marks of students in a class.
- (vi) Average intelligence of students in a class.
- (vii) Average size of collars.
- (viii) Average income of a lawyer.
- (ix) Average size of readymade garments.
- (x) Average size of agricultural holdings.
- (xi) Average change in prices.
- (xii) Average level of health.

118. Select the correct alternative.

- (a) Relationship between mean (m), geometric mean (g) and harmonic mean (h) is:

$$(i) g = \frac{m.h}{m+h} \quad (ii) g = \sqrt{m.h} \quad (iii) g = \frac{m+h}{2} \quad (iv) \text{None of the these.}$$

- (b) In a moderately skewed distribution, mode (M_o) can be calculated by:

$$(i) M_o = \frac{3\bar{X} - 2M_d}{2}, (ii) M_o = 3\bar{X} - 2M_d$$

$$(iii) M_o = 3\bar{X} - 3M_d, (iv) M_o = 3M_d - 2\bar{X}$$

- (c) Which of the following would be an appropriate average for determining the average size of readymade garments :

- (i) Arithmetic mean (ii) Median (iii) Mode (iv) Geometric mean

- (d) Most appropriate average to determine the size of oranges on a tree is:

- (i) Mode (ii) Median (iii) Mean (iv) None of the these.

- (e) Most appropriate measure for qualitative measurements is :

- (i) Mode (ii) Median (iii) Mean (iv) None of the these.

- (f) The most unstable measure of central tendency is :
 (i) Mean (ii) Median (iii) Mode (iv) None of the these.
- (g) The sum of deviations of observations is zero when measured from :
 (i) Median (ii) GM (iii) Mode (iv) Mean
- (h) The average, most affected by the extreme observations, is :
 (i) Mode (ii) Mean (iii) GM (iv) Median
- (i) The most stable average is :
 (i) Mode (ii) Mean (iii) Median (iv) GM
119. State whether the following statements are true or false :
 (i) \bar{X} can be calculated for a distribution with open ends.
 (ii) M_d is not affected by the extreme observations.
 (iii) \bar{X} is based on all the observations.
 (iv) $\bar{X} = M_o = M_d$, for a symmetrical distribution.
 (v) M_o can be calculated if class intervals are of unequal width.
 (vi) The class limits should be exclusive for the calculation of M_d and M_o .
120. Fill in the blanks :
 (i) is most suitable for measuring average rate of growth.
 (ii) or are used for averaging rates under certain conditions.
 (iii) or are the averages which can be calculated for a distribution with open ends.
 (iv) or are the averages used to study the pattern of a distribution.
 (v) or are the averages which can be calculated when the characteristics are not measurable.
 (vi) or or averages depend upon all the observations.
 (vii) The sum of squares of deviations is when taken from mean.
 (viii) The average which divides a distribution into two equal parts is
 (ix) M_d of a distribution is also equal to its quartile.
 (x) The point of intersection of the 'less than type' and 'more than type' ogives corresponds to
 (xi) The algebraic sum of deviations of 30 observations from a value 14 is 3. The mean of these observations is
121. Examine the validity of the following statements giving necessary proofs and reasons for your answer :
 (i) For a set of 50 observations X_i , $i = 1, 2 \dots 50$, $\sum_{i=1}^{50} (X_i - 10) = 90$, when $\bar{X} = 10$.
 (ii) Geometric mean of a given number of observations cannot be obtained if one of them is zero.
 (iii) The mean depth of water of a river is 130 cms, therefore, a man with a height of 165 cms can cross the river safely.
 (iv) For a wholesale manufacturer, interested in the type which is usually in demand, median is the most suitable average.

- (v) If $AM = 25$ and $HM = 9$, then $GM = 15$ for two positive values of a variable.
- (vi) For a set of 8 observations AM , GM and HM are 5.2, 6.3 and 7.1 respectively.
- (vii) If $2y - 6x = 6$ and mode of y is 66, then mode of x is 21.

2.17 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

1. (a) True (b) True (c) False (d) True (e) True
 2. (a) Average (b) Central Tendency (c) Median (d) Ten
(e) Inspection, grouping
-

2.18 SUGGESTED READINGS

Mario F. Triola, *Elementary Statistics*, Addison-Wesley January 2006.

David & Moae, *Introduction to the Practice of Statistics*, W.H. Freeman & Co. February 2005.

Allan & Blumon, *Elementary Statistics : A Step by Step Approach*. McGraw-Hill College, June 2003.

James T. McClave Terry Sincich, William Mendenhall, *Statistics*, Prentice Hall February 2005

Mark L. Berenson, David M. Revine, Tineothy C. Krehbiel, *Basic Business Statistics: Concepts & Applications*, Prentice Hall, May 2005.

LESSON

3

MATHEMATICAL MODEL

CONTENTS

- 3.0 Aims and Objectives
 - 3.1 Introduction
 - 3.2 Mathematics — The Language of Modelling
 - 3.3 Building a Mathematical Model
 - 3.4 Verifying and Refining a Model
 - 3.5 Variables and Parameters
 - 3.6 Continuous-in-Time vs. Discrete-in-Time Models
 - 3.7 Deterministic Model Example
 - 3.8 Probabilistic Models
 - 3.9 Let us Sum Up
 - 3.10 Lesson-end Activity
 - 3.11 Keywords
 - 3.12 Questions for Discussion
 - 3.13 Model Answers to Questions for Discussion
 - 3.14 Suggested Readings
-

3.0 AIMS AND OBJECTIVES

In the previous and preceding previous lessons we had an elaborative view on quantitative techniques and distinguishable statistical approaches. In this lesson we are going to discuss various models or rather we can say mathematical models which are extremely powerful because they usually enable predictions to be made about the system.

3.1 INTRODUCTION

Models in science come in different forms. A physical model that you probably are familiar with is an anatomically detailed model of the human body. Mathematical models are less commonly found in science classes, but they form the core of modern cosmology. Mathematical models are extremely powerful because they usually enable predictions to be made about a system. The predictions then provide a road map for further experimentation. Consequently, it is important for you to develop an appreciation for this type of model as you learn more about cosmology. Two sections of the activity develop mathematical models of direct relevance to cosmology and astronomy. The math skills required in the activity increase with each section, but nothing terribly advanced is required. A very common approach to the mathematical modeling of a physical system is to collect

a set of experimental data and then figure out a way to graph the data so that one gets a straight line. Once a straight line is obtained, it is possible to generalize the information contained in the straight line in terms of the powerful algebraic equation: You probably are familiar with this equation. In it y represents a value on the y -axis, x represents a value on the x -axis, m represents the slope of the straight line, and b represents the value of the intercept of the line on the y -axis. In all sections of this activity, your goal will be to analyze and then graph a set of data so that you obtain a straight line. Then you will derive the equation that describes the line, and use the equation to make predictions about the system. So relax and have fun with math!

$$y = mx + b$$

Mathematical modeling is the process of creating a mathematical representation of some phenomenon in order to gain a better understanding of that phenomenon. It is a process that attempts to match observation with symbolic statement. During the process of building a mathematical model, the model will decide what factors are relevant to the problem and what factors can be de-emphasized. Once a model has been developed and used to answer questions, it should be critically examined and often modified to obtain a more accurate reflection of the observed reality of that phenomenon. In this way, mathematical modeling is an evolving process; as new insight is gained, the process begins again as additional factors are considered. “Generally the success of a model depends on how easily it can be used and how accurate are its predictions.” (Edwards & Hamson, 1994, p. 3)

3.2 MATHEMATICS — THE LANGUAGE OF MODELLING

Like other languages, the essence of mathematics is the way it enables us to express, communicate, and reason about ideas and, especially, ideas about our world. The word “red” in English is important because it describes the color below. Without seeing this color one misses a great deal about the word “red.”

We are interested in using mathematics to talk about meaningful problems. For this reason, laboratory equipment like the Texas Instrument CBL that allows us to collect and record quantitative information about the real world, and sources like the United States Census are especially important to us.

Working with real problems requires the full power of mathematics — the ability to work with symbols, with graphics, and with numerical calculations. For this reason computer algebra systems like MathCad, Maple, Mathematica, and the CAS built into the TI-92 are an integral part of our tool kit. They give us powerful environments for doing mathematics. And together with a browser, like Netscape, some cables, and equipment like the TI-CBL they give us the ability to use the full power of mathematics with real data from the real world.

3.3 BUILDING A MATHEMATICAL MODEL

Building a mathematical model for your project can be challenging, yet interesting, task. A thorough understanding of the underlying scientific concepts is necessary and a mentor with expertise in your project topic is invaluable. It is also best to work as part of a team to provide more brainstorming power. In industry and engineering, it is common practice for a team of people to work together in building a model, with the individual team members bringing different areas of expertise to the project.

Although problems may require very different methods of solution, the following steps outline a general approach to the mathematical modeling process:

1. **Identify the problem,** define the terms in your problem, and **draw diagrams** where appropriate.

2. Begin with a **simple** model, **stating the assumptions** that you make as you focus on particular aspects of the phenomenon.
3. **Identify important variables and constants** and determine how they relate to each other.
4. **Develop the equation(s)** that express the relationships between the variables and constants.

Check Your Progress 3.1

1. What is the difference between physical model and mathematical model?
2. What are the different steps of mathematical modelling process?

Notes: (a) Write your answer in the space given below.
(b) Please go through the lesson sub-head thoroughly you will get your answers in it.
(c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

3.4 VERIFYING AND REFINING A MODEL

Once the model has been developed and applied to the problem, your resulting model solution must be analyzed and interpreted with respect to the problem. The interpretations and conclusions should be checked for accuracy by answering the following questions:

- Is the information produced reasonable?
- Are the assumptions made while developing the model reasonable?
- Are there any factors that were not considered that could affect the outcome?
- How do the results compare with real data, if available?

In answering these questions, you may need to modify your model. This refining process should continue until you obtain a model that agrees as closely as possible with the real world observations of the phenomenon that you have set out to model.

3.5 VARIABLES AND PARAMETERS

Mathematical models typically contain three distinct types of quantities: output variables, input variables, and parameters (constants). Output variables give the model solution. The choice of what to specify as input variables and what to specify as parameters is somewhat arbitrary and often model dependent. Input variables characterize a single physical problem while parameters determine the context or setting of the physical problem. For example, in modeling the decay of a single radioactive material, the initial amount of material and the time interval allowed for decay could be input variables, while the decay constant for the material could be a parameter. The output variable for this model is the amount of material remaining after the specified time interval.

3.6 CONTINUOUS-IN-TIME VS. DISCRETE-IN-TIME MODELS

Mathematical models of time dependent processes can be split into two categories depending on how the time variable is to be treated. A continuous-in-time mathematical model is based on a set of equations that are valid for any value of the time variable. A discrete-in-time mathematical model is designed to provide information about the state of the physical system only at a selected set of distinct times.

The solution of a continuous-in-time mathematical model provides information about the physical phenomenon at every time value. The solution of a discrete-in-time mathematical model provides information about the physical system at only a finite number of time values. Continuous-in-time models have two advantages over discrete-in-time models: (1) they provide information at all times and (2) they more clearly show the qualitative effects that can be expected when a parameter or an input variable is changed. On the other hand, discrete in time models have two advantages over continuous in time models: (1) they are less demanding with respect to skill level in algebra, trigonometry, calculus, differential equations, etc. and (2) they are better suited for implementation on a computer.

Some Examples of Mathematical Models

Population	Growth
A	Spring
A	Falling

Problem 1

Rotating all or part of a space station can create artificial gravity in the station. The resulting centrifugal force will be indistinguishable from gravitational force. Develop a mathematical model that will determine the rotational rate of the station as a function of the radius of the station (distance from the center of rotation) and the desired artificial gravitational force. Use this model to answer the question: What rotational rate is needed if the radius of the station is 150 m and Earth surface gravity is desired.

Problem 2

A stretch of Interstate 25 is being widened to accommodate increasing traffic going north and south. Unfortunately, the Department of Transportation is going to have to bring out the orange barrels and close all but one lane at the “big I” intersection. The department would like to have traffic move along as quickly as possible without additional accidents. What speed limit would provide for maximum, but safe, traffic flow?

3.7 DETERMINISTIC MODEL EXAMPLE

An example of a deterministic model is a calculation to determine the return on a 5-year investment with an annual interest rate of 7%, compounded monthly. The model is just the equation below:

$$F = P (1 + r/m)^{YM}$$

The inputs are the initial investment ($P = \$1000$), annual interest rate ($r = 7\% = 0.07$), the compounding period ($m = 12$ months), and the number of years ($Y = 5$).

Compound Interest Model

Present value, P	1000.00
Annual rate, r	0.07
Periods/Year, m	12
Years, Y	5
Future value, F	

One of the purposes of a model such as this is to make predictions and try “**What If?**” **scenarios**. You can change the inputs and recalculate the model and you’ll get a new answer. You might even want to plot a graph of the future value (F) vs. years (Y). In some cases, you may have a fixed interest rate, but what do you do if the interest rate is allowed to change? For this simple equation, you might only care to know a **worst/best case scenario**, where you calculate the future value based upon the lowest and highest interest rates that you might expect.

3.8 PROBABILISTIC MODELS

The toy roulette at the left is a pale model of a real roulette wheel. Real roulette wheels are usually found in casinos, surrounded by glitter and glitz. But this toy captures the essentials of roulette. Both the toy and real roulette wheels have 38 slots, numbered 1 through 36, 0, and 00. Two of the slots are colored green; 18 are colored red and 18 are colored black. Bettors often bet on red. If they wager \$1.00 on red then if the roulette ball lands in a red slot they win \$1.00 but if it lands in either a green slot or a black slot they lose \$1.00. Because there are 18 red slots out of a total of 38 slots the chances of winning this bet are 18/38 — considerably less than even. The casinos make up the rules and they make them up so that they make huge profits.

Gambling games like roulette are good models for many phenomena involving chance — for example, investing in the stock market. It is easier to analyze games involving a roulette wheel than investments involving the stock market but the same ideas are involved. In this section we will consider and compare two different strategies that a gambler might use playing roulette. The same kinds of strategies and considerations are involved with investments. The same tools that we develop here for roulette can be used by investors.



Suppose that you have \$10.00 and that you want to win an additional \$10.00. We will consider two different strategies.

- **The Flamboyant Strategy:**

You stride purposefully up to the wheel with a devil-may-care smile on your face. You bet your entire fortune of \$10.00 on one spin of the wheel. If the ball lands in a red slot then you win, pocket your winnings, and leave with \$20.00 and a genuine happy smile on your face. If the ball lands in a slot of a different color then you

smile bravely at everyone as if \$10.00 is mere chickenfeed and leave with empty pockets and feeling gloomy. **With the flamboyant strategy your chances of winning are 18/38 or roughly 0.4737.**

- **The Timid Strategy:**

With this strategy you approach the roulette table with obvious trepidation. After watching for a while and working up your courage, you bet \$1.00. When the ball falls in a slot you either win or lose \$1.00. Now you have either \$9.00 or \$11.00. You continue betting one dollar on each spin of the wheel until you either go broke or reach your goal of \$20.00.

Before continuing pause and think about these two strategies. Which of the two do you think gives you the best chance of winning? — or are your chances of winning the same whichever strategy you use?



One way to study the questions raised above is by trying the two strategies in real casinos, wagering your own real money. This approach has several advantages and several disadvantages. One advantage is that this approach is realistic. Real casinos are run by people who know how to make a profit. They are skilled at creating an atmosphere that is likely to encourage customers to bet and lose more than they might like. The lessons that you learn in a real casino are more likely to be real lessons than the ones you learn in a simulated casino like the one we use below. One disadvantage is that this approach can be very costly both in terms of money and time.

We take a different approach — using the CAS window to simulate playing with the second, or timid, strategy. We already know the chances of winning with the first, or flamboyant, strategy — 18/38, or roughly 0.4737.



Computer algebra systems like Maple, MathCad, Mathematica, or the CAS system in the TI-92 have a procedure that generates random numbers. For example, on the TI-92 the command rand(), produces a random number between zero and one. The screen below shows the results of executing this command seven times. Notice that it produced seven different random numbers.

F1	F2	F3	F4	F5	F6	
Algebra	Calc	Other	Prgm	ID	Clear	a-z
■ rand()						.332431
■ rand()						.650691
■ rand()						.509445
■ rand()						.159966
■ rand()						.189059
■ rand()						.58333
■ rand()						.902255
rand()						
MATH	MAT					FRNC 7/10



Using the random number generator in your CAS window, you can easily simulate one spin of a roulette with a procedure like the one shown below.

```
:spin()
:Prgm
:If rand()<18/38 Then
:Disp "win $1.00"
:Else
:Disp "lose $1.00"
:EndIf
:EndPrgm
```

Your CAS window has a program that is built on this basic idea and will simulate playing roulette using the timid strategy. Use this program to answer the questions below.



- Compare the timid strategy to the flamboyant strategy.
- Consider an intermediate strategy — betting \$2.00 on each spin of the wheel.
- Consider another, intermediate strategy — betting \$5.00 on each spin of the wheel.
- Some people enjoy gambling. If you play the flamboyant strategy then you spin the wheel just once. On the average how often would you spin the wheel with each of the strategies above.
- What conclusion can you draw from our work in this module regarding the advisability of diversifying your investments? Be careful. Your answer depends on your investment goals and your beliefs about whether stock prices are more likely to rise or to fall.

Check Your Progress 3.2

1. What is the difference between Variables and Parameters?
2. Give two applications of computer algebra system.

Notes: (a) Write your answer in the space given below.
(b) Please go through the lesson sub-head thoroughly you will get your answers in it.
(c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

3.9 LET US SUM UP

This course is about the essence of science — understanding the world in which we live. We use mathematics as a language to help us describe and understand our world. Because the purpose of Mathematical Modeling is to “talk about” our world, the most important part of this course are the applications — our mathematical discussions about real world

phenomena. In this first chapter we have looked at the following applications. Everything we have discussed above — the content of the course — the tools and the technology — would be useless without you. Indeed, without you there would be no purpose. The purpose of mathematical modeling is to enable people like you and me to learn about our world, to form mental pictures of how it works and how we can make it a bit better. Mathematical modeling requires your active participation — thinking, working with your computer algebra system, with old-fashioned paper and pencil, exploring the world with the TI-CBL, rubber bands, and TinkerToys, and exchanging ideas with friends and colleagues.

3.10 LESSON-END ACTIVITY

As we know that mathematical modelling is the process of creating a mathematical representation of some phenomenon. So constructing a mathematical model for your project can be a challenging, yet interesting task. Being a technician and a computer user you have to think of a system where mathematical modelling will be used. Like use of mathematical modelling in a National Stock Exchange.

3.11 KEYWORDS

Model

Time Models

Flamboyant Strategy

Timid Strategy

Parameters

3.12 QUESTIONS FOR DISCUSSION

1. Write True or False against each statement:

- (a) Mathematical Modelling is the process of creating a mathematical representation.
- (b) Mathematical models typically contain input and output variables and parameters.
- (c) Models only represent patterns found in graphs.
- (d) Mathematical Modelling is used to collect a set of experimental data and figure out to graph.

2. Distinguish between:

- (a) Variables and Parameters
- (b) Continuous-in-Time and Discrete-in-Time Mathematical Model
- (c) The Flamboyant Strategy and The Timid Strategy
- (d) Probabilistic Model and Deterministic Model
- (e) Mathematics and Mathematical Modelling

3. Write short notes on:

- (a) Model
- (b) Building a Mathematical Model
- (c) Time Mathematical Model

- (d) Compound Interest Model
- (e) Probabilistic Models Strategy

4. Fill in the blanks:

- (a) A good model should the essential character of the model to be analysed.
- (b) Building Model can be yet task.
- (c) model is based on a set of equations.
- (d) The best example of probabilistic model is
- (e) is based other than flamboyant and timid strategy.

3.13 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

- 1. (a) True (b) True (c) False (d) True
- 4. (a) Capture (b) Challenging, interesting (c) Continuous-in-Time
(d) Gambling games (e) CAS window

3.14 SUGGESTED READINGS

M.P. Williams, *Model Building in Mathematical Programming*, Wiley

Bruno Poizat, Moses Klein, *A Course in Model Theory: An Introduction to Contemporary Mathematical Logic*, Springer Verlag, May 2000

N.N. Bogolubov, Jr, A.S. Shunovsly, B.I. Sadovnikov, A.S. Shumovskii, *Mathematical Methods of Statistical Mechanics of Model System CRC*, Pr I LIC

A.H. Lingstone, *Mathematical Logic: An Introduction to Model Theory*, Kluwer Academic (Pub)

Goran Pjordevic, J.Wess, Lyubisa Nesic, *Mathematical, Theoretical and Phenomenological Challenges*, World Scientific Pub Co. Inc. (March 2005)

LESSON

4

LINEAR PROGRAMMING: GRAPHICAL METHOD

CONTENTS

- 4.0 Aims and Objectives
- 4.1 Introduction
- 4.2 Essentials of Linear Programming Model
- 4.3 Properties of Linear Programming Model
- 4.4 Formulation of Linear Programming
- 4.5 General Linear Programming Model
- 4.6 Maximization & Minimization Models
- 4.7 Graphical Method
- 4.8 Solving Linear Programming Graphically Using Computer
- 4.9 Summary of Graphical Method
- 4.10 Unbounded LP Problem
- 4.11 Let us Sum Up
- 4.12 Lesson-end Activity
- 4.13 Keywords
- 4.14 Questions for Discussion
- 4.15 Terminal Questions
- 4.16 Model Answers to Questions for Discussion
- 4.17 Suggested Readings

4.0 AIMS AND OBJECTIVES

In this unit we have talked about Quantitative Techniques and the Measurement of Mean, Median and Mode and the various Mathematical Models and now we will talk about linear programming and in this lesson we will learn the graphical method of linear programming.

4.1 INTRODUCTION

Linear programming is a widely used mathematical modeling technique to determine the optimum allocation of scarce resources among competing demands. Resources typically include raw materials, manpower, machinery, time, money and space. The technique is very powerful and found especially useful because of its application to many different types of real business problems in areas like finance, production, sales and distribution, personnel, marketing and many more areas of management. As its name implies, the

linear programming model consists of linear objectives and linear constraints, which means that the variables in a model have a proportionate relationship. For example, an increase in manpower resource will result in an increase in work output.

4.2 ESSENTIALS OF LINEAR PROGRAMMING MODEL

For a given problem situation, there are certain essential conditions that need to be solved by using linear programming.

1. Limited resources : limited number of labour, material equipment and finance
2. Objective : refers to the aim to optimize (maximize the profits or minimize the costs).
3. Linearity : increase in labour input will have a proportionate increase in output.
4. Homogeneity : the products, workers' efficiency, and machines are assumed to be identical.
5. Divisibility : it is assumed that resources and products can be divided into fractions. (in case the fractions are not possible, like production of one-third of a computer, a modification of linear programming called integer programming can be used).

4.3 PROPERTIES OF LINEAR PROGRAMMING MODEL

The following properties form the linear programming model:

1. Relationship among decision variables must be linear in nature.
2. A model must have an objective function.
3. Resource constraints are essential.
4. A model must have a non-negativity constraint.

4.4 FORMULATION OF LINEAR PROGRAMMING

Formulation of linear programming is the representation of problem situation in a mathematical form. It involves well defined decision variables, with an objective function and set of constraints.

Objective function

The objective of the problem is identified and converted into a suitable objective function. The objective function represents the aim or goal of the system (i.e., decision variables) which has to be determined from the problem. Generally, the objective in most cases will be either to maximize resources or profits or, to minimize the cost or time.

For example, assume that a furniture manufacturer produces tables and chairs. If the manufacturer wants to maximize his profits, he has to determine the optimal quantity of tables and chairs to be produced.

Let x_1 = Optimal production of tables
 p_1 = Profit from each table sold

x_2 = Optimal production of chairs
 p_2 = Profit from each chair sold.

Hence, Total profit from tables = $p_1 x_1$

Total profit from chairs = $p_2 x_2$

The objective function is formulated as below,

$$\text{Maximize } Z \text{ or } Z_{\max} = p_1 x_1 + p_2 x_2$$

Constraints

When the availability of resources are in surplus, there will be no problem in making decisions. But in real life, organizations normally have scarce resources within which the job has to be performed in the most effective way. Therefore, problem situations are within confined limits in which the optimal solution to the problem must be found.

Considering the previous example of furniture manufacturer, let w be the amount of wood available to produce tables and chairs. Each unit of table consumes w_1 unit of wood and each unit of chair consumes w_2 units of wood.

For the constraint of raw material availability, the mathematical expression is,

$$w_1 x_1 + w_2 x_2 \leq w$$

In addition to raw material, if other resources such as labour, machinery and time are also considered as constraint equations.

Non-negativity constraint

Negative values of physical quantities are impossible, like producing negative number of chairs, tables, etc., so it is necessary to include the element of non-negativity as a constraint i.e., $x_1, x_2 \geq 0$

4.5 GENERAL LINEAR PROGRAMMING MODEL

A general representation of LP model is given as follows:

$$\text{Maximize or Minimize, } Z = p_1 x_1 + p_2 x_2 + \dots + p_n x_n$$

Subject to constraints,

$$w_{11} x_1 + w_{12} x_2 + \dots + w_{1n} x_n \leq \text{ or } = \text{ or } \geq w_1 \dots \quad (i)$$

$$w_{21} x_1 + w_{22} x_2 + \dots + w_{2n} x_n \leq \text{ or } = \text{ or } \geq w_2 \dots \quad (ii)$$

.

.

.

$$w_{m1} x_1 + w_{m2} x_2 + \dots + w_{mn} x_n \leq \text{ or } = \text{ or } \geq w_m \dots \quad (iii)$$

Non-negativity constraint,

$$x_i \geq 0 \text{ (where } i = 1, 2, 3, \dots, n)$$

Check Your Progress 4.1

1. What are the essentials of LP Model?
2. Why linear programming is used?

Contd...

- Notes:**
- Write your answer in the space given below.
 - Please go through the lesson sub-head thoroughly you will get your answers in it.
 - This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-
-

4.6 MAXIMIZATION & MINIMIZATION MODELS

Example 1: A biscuit manufacturing company plans to produce two types of biscuits, one with a round shape and another with a square shape. The following resources are used in manufacturing the biscuits,

- Raw material, of which daily availability is 150 kg.
- Machinery, of which daily availability is 25 machine hours.
- Labour, of which daily availability is 40 man-hours.

The resources used are shown in Table 1. If the unit profit of round and square biscuits is Rs 3.00 and Rs 2.00 respectively, how many round and square biscuits should be produced to maximize total profit ?

Table 4.1: Resources Used

Resources	Requirement/Unit		Daily availability
	Round	Square	
Raw Material	100	115	1500 grams
Machine	10	12	720 minutes
Manpower	3	2	240 minutes

Solution:

Key Decision: To determine the number of round and square biscuits to be produced.

Decision Variables:

Let x_1 be the number of round biscuits to be produced daily, and

x_2 be the number of square biscuits to be produced daily

Objective function: It is given that the profit on each unit of round biscuits is Rs 3.00 and of square biscuits is Rs. 2.00. The objective is to maximize profits, therefore, the total profit will be given by the equation,

$$Z_{\max} = 3x_1 + 2x_2$$

Constraints: Now, the manufacturing process is imposed by a constraint with the limited availability of raw material. For the production of round biscuits, $100x_1$ of raw material is used daily and for the production of square biscuits, $115x_2$ of raw material is used daily. It is given that the total availability of raw material per day is 1500 grams.

Therefore, the constraint for raw material is,

$$100x_1 + 115x_2 \leq 1500$$

Similarly, the constraint for machine hours is,

$$10x_1 + 12x_2 \leq 720$$

and for the manpower is.

$$3x_1 + 2x_2 \leq 240$$

Since the resources are to be used within or below the daily available level, inequality sign of less than or equal sign (\leq) is used. Further, we cannot produce negative number of units of biscuits which is a non-negative constraint expressed as,

$$x_1 \geq 0 \text{ and } x_2 \geq 0$$

Thus, the linear programming model for the given problem is,

$$\text{Maximize } Z = 3x_1 + 2x_2$$

Subject to constraints,

$$100x_1 + 115x_2 \leq 1500 \quad \dots \dots \dots \text{(i)}$$

$$10x_1 + 12x_2 \leq 720 \quad \dots \dots \dots \text{(ii)}$$

$$3x_1 + 2x_2 \leq 240 \quad \dots \dots \dots \text{(iii)}$$

where $x_1 \geq 0, x_2 \geq 0$

Example 2: Rahul Ads, an advertising company is planning a promotional campaign for the client's product, i.e., sunglasses. The client is willing to spend Rs. 5 lakhs. It was decided to limit the campaign media to a weekly magazine, a daily newspaper and TV advertisement. The product is targeted at middle-aged men and women, and the following data was collected (Table 4.2).

Table 4.2: Data Collected

Campaign media	Cost per advertisement (Rs.)	Expected viewers
Weekly Magazine	30,000	1,15,000
Daily Newspaper	45,000	2,05,000
TV Advetisement	1,25,000	7,00,000

The client is interested to spend only Rs. 1 lakh on the ads in the weekly magazine which is expecting a viewership of a minimum of 21 lakh people in the case of the television advertising. Maximize the viewers to the advertisements.

Solution:

Key Decision: To determine number of advertisements on weekly magazine, daily newspaper and TV.

Let x_1 be the number of weekly magazine advertisements.

x , be the number of daily newspaper advertisements.

x_3 be the number of TV advertisements.

Objective function: The objective is to maximize the number of viewers through all media. The total viewers will be given by the equation,

$$Z_{\max} = 115000x_1 + 205000x_2 + 700000x_3$$

Constraints: Firstly, the client is willing to spend Rs. 500000 on all media,

$$30000x_1 + 45000x_2 + 125000x_3 \leq 500000$$

or

$$30x_1 + 45x_2 + 125x_3 \leq 500 \quad \dots \dots \dots \text{(i)}$$

Secondly, a minimum of 2100000 people should view the television advertising,

$$700000x \geq 2100000$$

or

Lastly, the client is interested to pay only Rs. 100000 in weekly magazine advertising,

$$30000x_1 \leq 100000$$

or

$$3x_1 \leq 10 \quad \dots \dots \dots \text{(iii)}$$

Summarizing the LP model for the given problem,

$$\text{Maximize } Z = 115000x_1 + 205000x_2 + 700000x_3$$

Subject to constraints,

$$30x_1 + 45x_2 + 125x_3 \leq 500 \quad \dots \dots \dots \text{(i)}$$

$$3x_1 \leq 10 \quad \dots \dots \dots \text{(iii)}$$

where $x_1, x_2, x_3 \geq 0$

Example 3: The data given in Table 4.3 represents the shipping cost (in Rs.) per unit for shipping from each warehouse to each distribution centre. The supply and demand data of each warehouse and distribution centre is given. Determine how many units should be shipped from each warehouse to each centre in order to minimize the overall transportation cost.

Table 4.3: Data Shows Shipping Cost from Warehouse to Distribution

Distribution Centre				
Warehouse	1	2	3	Supply
1	9	10	11	150
2	4	6	8	250
Demand	150	100	150	400

Solution:

Decision Variables

Let x_{ij} be the number of units to be shipped from warehouse i to distribution centre j .

x_{11} be the number of units to be shipped from warehouse 1 to distribution centre 1.

x_{12} be the number of units to be shipped from warehouse 1 to distribution centre 2.

x_{13} be the number of units to be shipped from warehouse 1 to distribution centre 3.

x_{21} be the number of units to be shipped from warehouse 2 to distribution centre 1.

x_{22} be the number of units to be shipped from warehouse 2 to distribution centre 2.

x_{23} be the number of units to be shipped from warehouse 2 to distribution centre 3.

Objective Function: The Table 4.3 shows the transportation cost from each warehouse to each distribution centre. Therefore $9x_{11}$ represents the total cost of shipping x_{11} units from warehouse 1 to distribution centre 1. The objective function is to minimize the transportation cost. Therefore, the objective function is,

$$\text{Minimize } Z = 9x_{11} + 10x_{12} + 11x_{13} + 4x_{21} + 6x_{22} + 8x_{23}$$

Constraints: The supply and demand constraints to ship the units from warehouses are, to ship the units and distribution centres must receive the shipped units. Since the given table is a 2×3 matrix we have a total 5 constraints apart from the non-negativity constraint. The constraints are as follows.

Linear Programming: Graphical Method

where $x_{ij} \geq 0$ ($i=1,2$, and $j=1,2,3$)

Thus the LP model for the given transportation problem is summarized as,

$$\text{Minimize } Z = 9x_{11} + 10x_{12} + 11x_{13} + 4x_{21} + 6x_{22} + 8x_{23}$$

Subject to constraints,

where $x_{ij} > 0$ ($i=1,2$, and $j = 1,2,3$)

Example 4: Sivakumar & Co., manufactures two types of T-shirts, one with collar and another without collar. Each T-shirt with collar yields a profit of Rs. 20, while each T-shirt without collar yields Rs. 30. Shirt with collar requires 15 minutes of cutting and 25 minutes of stitching. Shirt without collar requires 10 minutes of cutting and 20 minutes of stitching. The full shift time is available for cutting in an 8 hour shift, but only 6 hours are available for stitching. Formulate the problem as an LP model to maximize the profit.

Solution:

Key decision: To determine the number of T-shirts with collar and without collar to be manufactured.

Decision variables:

Let x be the number of T-shirts with collar.

x_2 be the number of T-shirts without collar

Objective Function:

$$Z_{\max} = 20x_1 + 30x_2$$

Constraints:

$$25x_1 + 20x_2 \leq 6 \times 60 \text{ (Stitching)} \quad \dots \dots \dots \text{(ii)}$$

Non-negativity constraints:

$$x_1 \geq 0, x_2 \geq 0$$

The linear programming model is,

$$Z_{\max} = 20x_1 + 30x_2$$

Subject to constraints,

$$15x_s + 10x_u \leq 480 \quad \dots \dots \dots \text{(i)}$$

$$25x_1 + 20x_2 \leq 360 \quad \dots \dots \dots \text{(ii)}$$

where $x_1, x_2 \geq 0$

Example 5: An agricultural urea company must daily produce 500 kg of a mixture consisting of ingredients x_1 , x_2 and x_3 . Ingredient x_1 costs Rs. 30 per kg, x_2 Rs. 50 per kg and x_3 Rs. 20 per kg. Due to raw material constraint, not more than 100 kg of x_1 , 70 kg of x_2 and 45 kg of x_3 must be used. Determine how much of each ingredient should be used if the company wants to minimize the cost.

Solution:

Let x_1 be the kg of ingredient x_1 to be used

x_2 be the kg of ingredient x_2 to be used

x_3 be the kg of ingredient x_3 to be used

The objective is to minimize the cost,

$$\text{Minimize } Z = 30x_1 + 50x_2 + 20x_3$$

Subject to constraints,

$x_1 \leq 100$ (max. use of x_1)(ii)

$x_2 \leq 70$ (max. use of x_2)(iii)

$x_2 \leq 45$ (max. use of x_2)(iv)

where $x_1, x_2, x_3 \geq 0$ (non-negativity)

Example 6: Chandru Bag Company produces two types of school bags: deluxe and ordinary. If the company is producing only ordinary bags, it can make a total of 200 ordinary bags a day. Deluxe bag requires twice as much labour and time as an ordinary type. The demand for deluxe bag and ordinary bag are 75 and 100 bags per day respectively. The deluxe bag yields a profit of Rs 12.00 per bag and ordinary bag yields a profit of Rs. 7.00 per bag. Formulate the problem as LP model.

Solution:

Let x_1 be deluxe bags to be produced per day

x_2 be ordinary bags to be produced per day

Objective function: The objective is to maximize the profit. Deluxe bag yields a profit of Rs. 12.00 per bag and ordinary bag yields a profit of Rs. 7.00 per bag.

$$\text{Maximize } Z = 12x_1 + 7x_2$$

Constraints: There are two constraints in the problem, the "number of bags" constraint and "demand" constraint. It is given that the deluxe bag takes twice as much time of ordinary bag and if only ordinary bags alone are produced, the company can make 200 bags.

The constraint is,

$$2x_1 + x_2 \leq 200$$

The demand for the deluxe bag is 75 bags and ordinary bag is 100 bags

The constraints are,

$$x_1 \leq 75$$

$$x_2 \leq 100$$

and the non-negativity constraint is,

$$x_+ > 0, x_- > 0$$

The LP formulation is

Linear Programming: Graphical Method

Maximize, $Z = 12x_1 + 7x_2$

Subject to constraints,

$$2x_1 + x_2 \leq 200 \quad \dots \dots \dots \text{(i)}$$

$$x_1 \leq 75 \quad \dots \dots \dots \text{(ii)}$$

$$x_2 \leq 100 \quad \dots \dots \dots \text{(iii)}$$

where $x_1, x_2 \geq 0$

Example 7: Geetha Perfume Company produces both perfumes and body spray from two flower extracts F_1 and F_2 . The following data is provided:

Table 4.4: Data Collected

Litres of Extract			
	Perfume	Body Spray	Daily Availability (litres)
Flower Extract, F ₁	8	4	20
Flower Extract, F ₂	2	3	8
Profit Per litre (Rs.)	7	5	

The maximum daily demand of body spray is 20 bottles of 100 ml each. A market survey indicates that the daily demand of body spray cannot exceed that of perfume by more than 2 litres. The company wants to find out the optimal mix of perfume and body spray that maximizes the total daily profit. Formulate the problem as a linear programming model.

Solution:

Let x_1 be the litres of perfume produced daily.

x , be the litres of body spray produced daily.

Objective function: The company wants to increase the profit by optimal product mix

$$Z_{\max} = 7x_1 + 5x_2$$

Constraints: The total availability of flower extract F_1 and flower extract F_2 are 20 and 8 litres respectively. The sum of flower extract F_1 used for perfume and body spray must not exceed 20 litres. Similarly, flower extract F_2 must not exceed 8 litres daily.

The constraints are,

$$8x_1 + 4x_2 \leq 20 \text{ (Flower extract F)}$$

$$2x_1 + 3x_2 \leq 8 \quad (\text{Flower extract } F_2)$$

The daily demand of body spray x_2 is limited to 20 bottles of 100ml each (i.e., $20 \times 100 = 2000$ ml = 2 litres)

Therefore, $x_2 \leq 2$

Again, there is an additional restriction, that the difference between the daily production of perfume and body spray , $x_2 - x_1$, does not exceed 2 litres, which is expressed as

$$x_3 - x_1 \leq 2$$

(or)

$$-x_1 + x_2 \leq 2.$$

The model for Geetha perfumes company is,

$$\text{Maximize } Z = 7x_1 + 5x_2$$

Subject to constraints,

where $x_1, x_2 \geq 0$

Feasible Solution: Any values of x_1 and x_2 that satisfy all the constraints of the model constitute a feasible solution. For example, in the above problem if the values of $x_1 = 2$ and $x_2 = 1$ are substituted in the constraint equation, we get

$$(i) \quad 8(2) + 4(1) \leq 20$$

20 ≤ 20

$$(ii) \quad 2(2) + 3(1) \leq 8$$

7 ≤ 8

(iii) $-2 + 1 \leq 2$

$$-1 \leq 2$$

(iv) $1 \leq 2$

All the above constraints (including non-negativity constraint) are satisfied. The objective function for these values of $x_1 = 2$ and $x_2 = 1$, are

$$Z_{\max} = 7(2) + 5(1) \\ \equiv 14 + 5 \equiv \text{Rs. } 19.00$$

As said earlier, all the values that do not violate the constraint equations are feasible solutions. But, the problem is to find out the values of x_1 and x_2 to obtain the optimum feasible solution that maximizes the profit. These optimum values of x_1 and x_2 can be found by using the Graphical Method or by Simplex Method. (The above problem is solved using graphical method shown on page number 117).

4.7 GRAPHICAL METHOD

Linear programming problems with two variables can be represented and solved graphically with ease. Though in real-life, the two variable problems are practiced very little, the interpretation of this method will help to understand the simplex method. The solution method of solving the problem through graphical method is discussed with an example given below.

Example 8: A company manufactures two types of boxes, corrugated and ordinary cartons. The boxes undergo two major processes: cutting and pinning operations. The profits per unit are Rs. 6 and Rs. 4 respectively. Each corrugated box requires 2 minutes for cutting and 3 minutes for pinning operation, whereas each carton box requires 2 minutes for cutting and 1 minute for pinning. The available operating time is 120 minutes and 60 minutes for cutting and pinning machines. Determine the optimum quantities of the two boxes to maximize the profits.

Solution:

Key Decision: To determine how many (number of) corrugated and carton boxes are to be manufactured.

Decision variables:

Let x be the number of corrugated boxes to be manufactured.

x_1 be the number of carton boxes to be manufactured.

Objective Function: The objective is to maximize the profits. Given profits on corrugated box and carton box are Rs. 6 and Rs. 4 respectively.

The objective function is,

$$Z_{\max} = 6x_1 + 4x_2$$

Constraints: The available machine-hours for each machine and the time consumed by each product are given.

Therefore, the constraints are,

$$2x_1 + x_2 \leq 60 \quad \dots \dots \dots \text{(ii)}$$

where $x_1, x_2 \geq 0$

Graphical Solution: As a first step, the inequality constraints are removed by replacing ‘equal to’ sign to give the following equations:

Find the co-ordinates of the lines by substituting $x_1 = 0$ and $x_2 = 0$ in each equation. In equation (1), put $x_1 = 0$ to get x_2 , and vice versa

$$2x_1 + 3x_2 = 120$$

$$2(0) + 3x_2 = 120, \quad x_2 = 40$$

Similarly, put $x_2 = 0$,

$$2x_1 + 3x_2 = 120$$

$$2x_1 + 3(0) = 120, \quad x_1 = 60$$

The line $2x_1 + 3x_2 = 120$ passes through co-ordinates (0, 40) (60, 0).

The line $2x_1 + x_2 = 60$ passes through co-ordinates $(0,60)(30,0)$.

The lines are drawn on a graph with horizontal and vertical axis representing boxes x_1 and x_2 , respectively. Figure 4.1 shows the first line plotted.

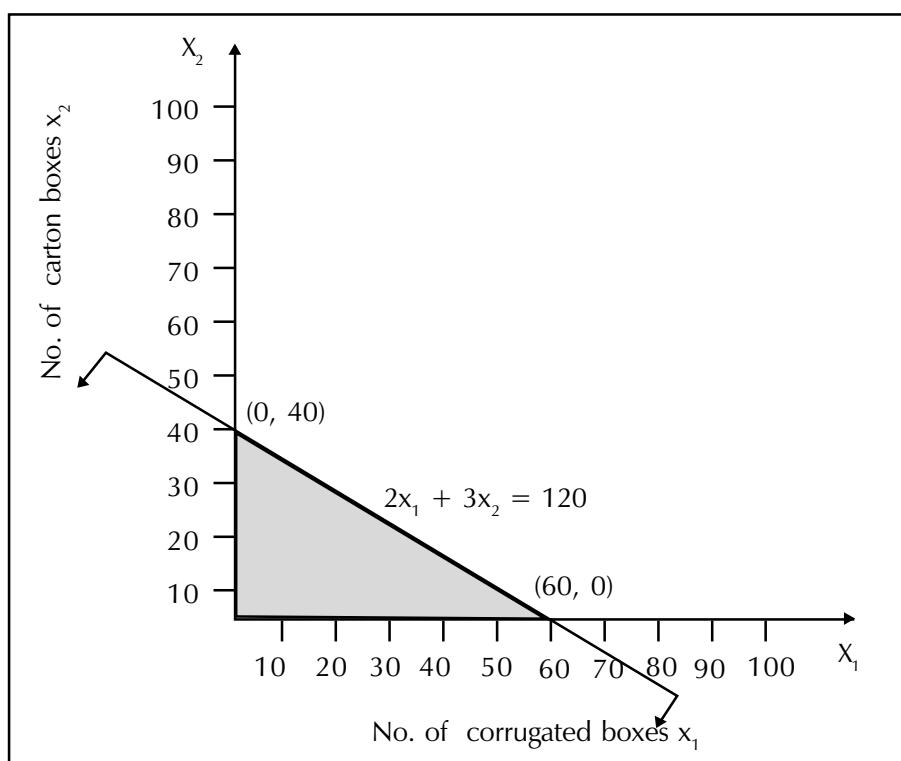


Figure 4.1: Graph Considering First Constraint

The inequality constraint of the first line is (less than or equal to) \leq type which means the feasible solution zone lies towards the origin. The no shaded portion can be seen is the feasible area shown in Figure 4.2 (Note: If the constraint type is \geq then the solution zone area lies away from the origin in the opposite direction). Now the second constraints line is drawn.

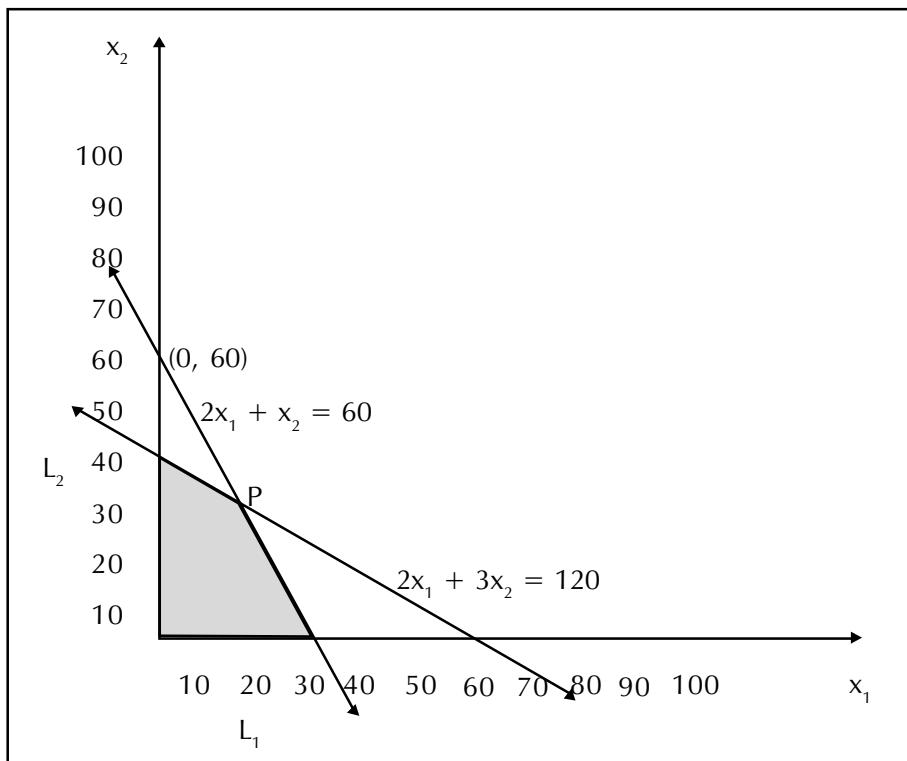


Figure 4.2: Graph Showing Feasible Area

When the second constraint is drawn, you may notice that a portion of feasible area is cut. This indicates that while considering both the constraints, the feasible region gets reduced further. Now any point in the shaded portion will satisfy the constraint equations. For example, let the solution point be (15,20) which lies in the feasible region.

If the points are substituted in all the equations, it should satisfy the conditions.

$$2x_1 + 3x_2 \leq 120 = 30 + 60 \leq 120 = 90 \leq 120$$

$$2x_1 + x_2 \leq 60 = 30 + 20 \leq 60 = 50 \leq 60$$

Now, the objective is to maximize the profit. The point that lies at the furthermost point of the feasible area will give the maximum profit. To locate the point, we need to plot the objective function (profit) line.

Equate the objective function for any specific profit value Z,

Consider a Z-value of 60, i.e.,

$$6x_1 + 4x_2 = 60$$

Substituting $x_1 = 0$, we get $x_2 = 15$ and

if $x_2 = 0$, then $x_1 = 10$

Therefore, the co-ordinates for the objective function line are (0,15), (10,0) as indicated by dotted line L_1 in Figure 4.2. The objective function line contains all possible combinations of values of x_1 and x_2 .

The line L_1 does not give the maximum profit because the furthermost point of the feasible area lies above the line L_1 . Move the line (parallel to line L_1) away from the origin to locate the furthermost point. The point P, is the furthermost point, since no area is seen further. Take the corresponding values of x_1 and x_2 from point P, which is 15 and 30 respectively, and are the optimum feasible values of x_1 and x_2 .

Therefore, we conclude that to maximize profit, 15 numbers of corrugated boxes and 30 numbers of carton boxes should be produced to get a maximum profit. Substituting $x_1 = 15$ and $x_2 = 30$ in objective function, we get

$$\begin{aligned} Z_{\max} &= 6x_1 + 4x_2 \\ &= 6(15) + 4(30) \end{aligned}$$

Maximum profit : Rs. 210.00

4.8 SOLVING LINEAR PROGRAMMING GRAPHICALLY USING COMPUTER

The above problem is solved using computer with the help of TORA. Open the TORA package and select LINEAR PROGRAMMING option. Then press Go to Input and enter the input data as given in the input screen shown below, in Figure 4.3.

INPUT GRID - LINEAR PROGRAMMING				
	x1	x2	Enter <, >, or =	R.H.S.
Var. Name	Corrugated	Carton		
Maximize	6.	4.		
Constr 1	2.	3.	<=	120.
Constr 2	2.	1.	<=	60.
Lower Bound	0.	0.		
Upper Bound	infinity	infinity		
Unrestr'd (y/n)?	n	n		

Figure 4.3: Linear Programming, TORA Package (Input Screen)

Now, go to Solve Menu and click Graphical in the 'solve problem' options. Then click Graphical , and then press Go to Output . The output screen is displayed with the graph grid on the right hand side and equations in the left hand side. To plot the graphs one by one, click the first constraint equation. Now the line for the first constraint is drawn connecting the points (40, 60). Now, click the second equation to draw the second line on the graph. You can notice that a portion of the graph is cut while the second constraint is also taken into consideration. This means the feasible area is reduced further. Click on the objective function equation. The objective function line locates the furthestmost point (maximization) in the feasible area which is (15,30) shown in Figure 4.4 below.

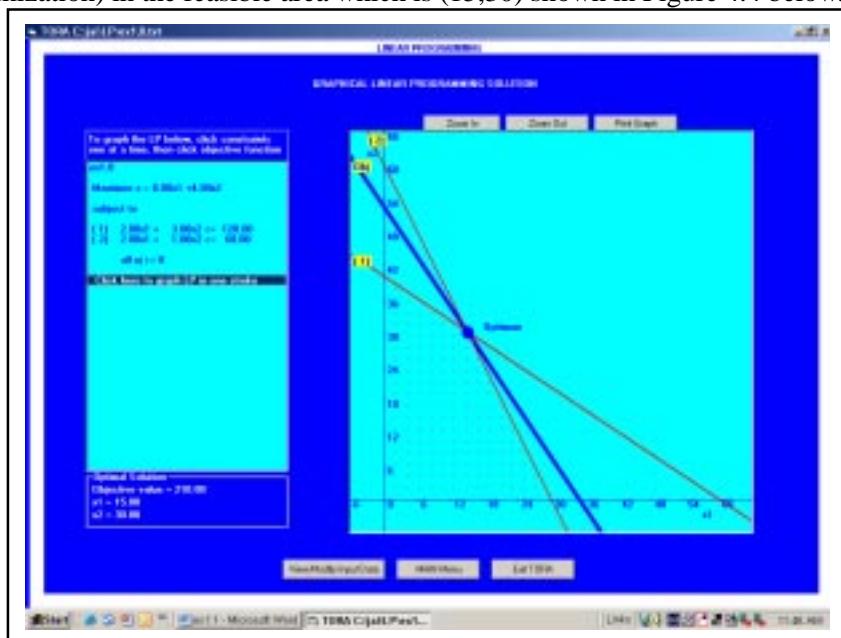


Figure 4.4: Graph Showing Feasible Area

Example 9: A soft drink manufacturing company has 300 ml and 150 ml canned cola as its products with profit margin of Rs. 4 and Rs. 2 per unit respectively. Both the products have to undergo process in three types of machine. The following Table 4.5, indicates the time required on each machine and the available machine-hours per week.

Table 4.5: Available Data

Requirement	Cola 300 ml	Cola 150 ml	Available machine-hours per week
Machine 1	3	2	300
Machine 2	2	4	480
Machine 3	5	7	560

Formulate the linear programming problem specifying the product mix which will maximize the profits within the limited resources. Also solve the problem using computer.

Solution: Let x_1 be the number of units of 300 ml cola and x_2 be the number of units of 150 ml cola to be produced respectively. Formulating the given problem, we get

Objective function:

$$Z_{\max} = 4x_1 + 2x_2$$

Subject to constraints.

$$2x_1 + 4x_2 \leq 480 \quad \dots \dots \dots \text{(ii)}$$

$$5x_s + 7x_u \leq 560 \quad \dots \dots \dots \text{(iii)}$$

where $x_1, x_2 \geq 0$

The inequalities are removed to give the following equations:

$$2x_1 + 4x_2 = 480 \quad (\text{v})$$

$$5x_1 + 7x_2 \equiv 560 \quad (\text{vi})$$

Find the co-ordinates of lines by substituting $x = 0$ to find y , and $y = 0$ to find x .

Therefore

Line $3x_1 + 2x_2 = 300$ passes through $(0, 150)$, $(100, 0)$

Line $2x_1 + 4x_2 = 480$ passes through $(0, 120)$, $(240, 0)$.

Line $5x_1 + 7x_2 = 650$ passes through $(0, 80)$, $(112, 0)$.

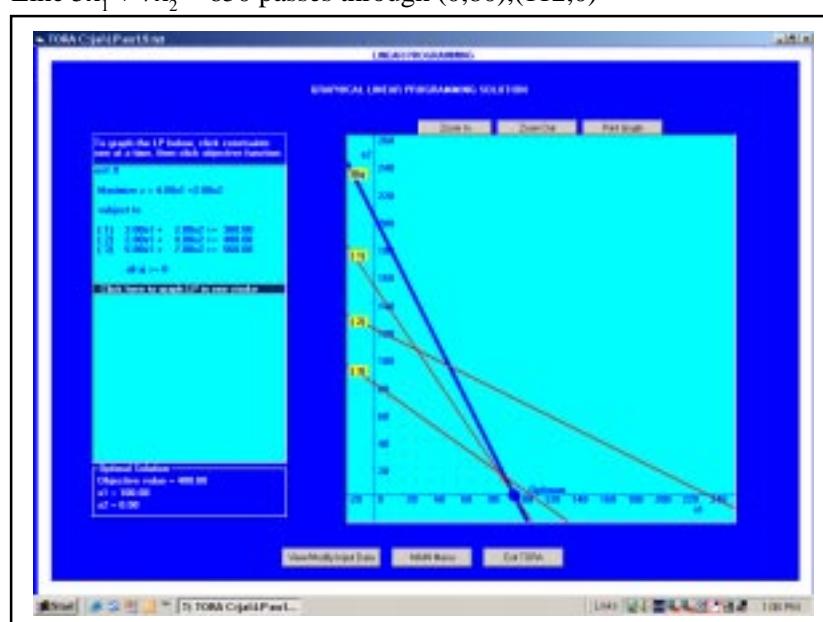


Figure 4.5: Graphical Presentation of lines (TOBA, Output Screen)

For objective function,

The Line $4x_1 + 2x_2 = 0$ passes through $(-10, 20), (10, -20)$

Plot the lines on the graph as shown in the computer output Figure 4.5.

The objective is to maximize the profit. Move the objective function line away from the origin by drawing parallel lines. The line that touches the furthermost point of the feasible area is $(100, 0)$. Therefore, the values of x_1 and x_2 are 100 and 0 respectively.

$$\begin{aligned}\text{Maximum Profit, } Z_{\max} &= 4x_1 + 2x_2 \\ &= 4(100) + 2(0) \\ &= \text{Rs. } 400.00\end{aligned}$$

Example 10: Solve the following LPP by graphical method.

$$\text{Minimize } Z = 18x_1 + 12x_2$$

Subject to constraints,

where $x_1, x_2 \geq 0$

Solution:

The inequality constraints are removed to give the equations,

The equation lines pass through the co-ordinates as follows:

For constraints,

$2x_1 + 4x_2 = 60$ passes through $(0,15)$, $(30,0)$.

$3x_1 + x_2 = 30$ passes through $(0,30)$, $(10,0)$.

$8x_1 + 4x_2 = 120$ passes through $(0,30)$, $(15,0)$.

The objective function,

$18x_1 + 12x_2 = 0$ passes through $(-10, 15)$, $(10, -15)$.

Plot the lines on the graph as shown in Figure 4.6.

Here the objective is minimization. Move the objective function line and locate a point in the feasible region which is nearest to the origin, i.e., the shortest distance from the origin. Locate the point P, which lies on the x – axis. The co-ordinates of the point P are (15,0) or $x_1 = 15$ and $x_2 = 0$.

The minimum value of Z

$$\begin{aligned} Z_{\min} &= 18 x_1 + 12x_2 \\ &= 18(15) + 12(0) \\ &\equiv \text{Rs. } 270.00 \end{aligned}$$

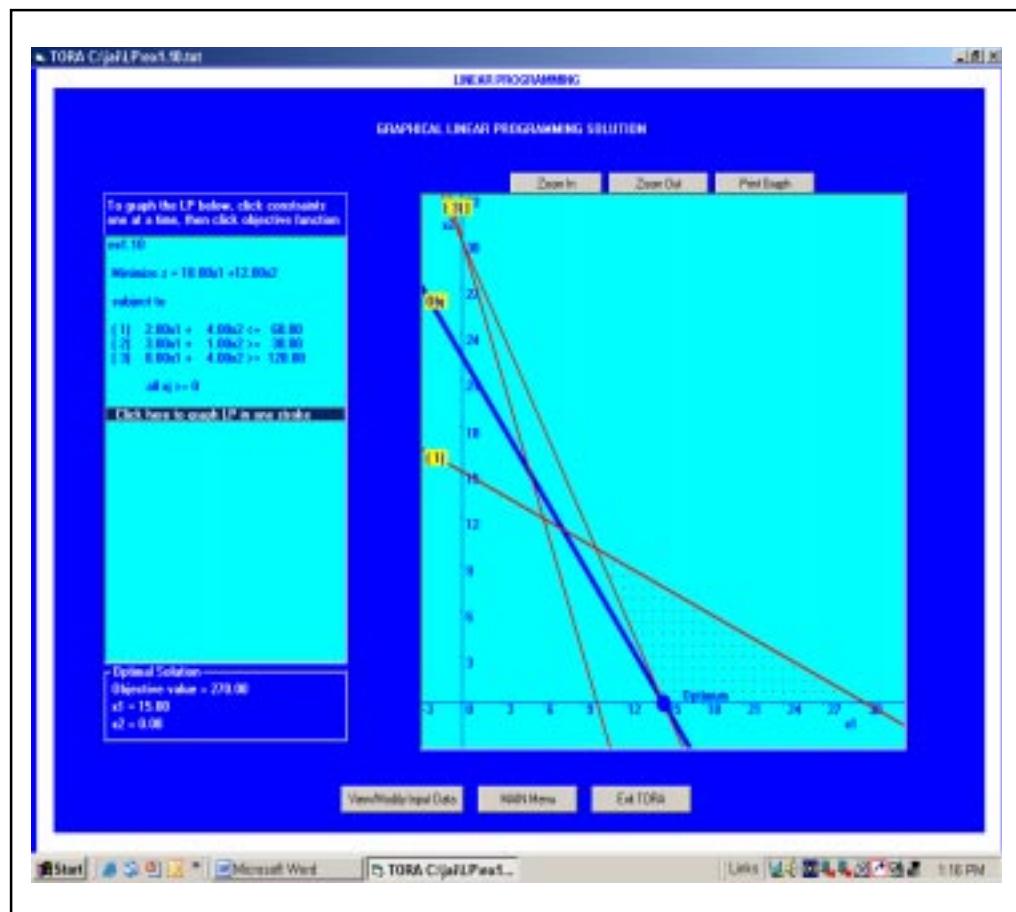


Figure 4.6: Graphical Presentation (Output Screen, TORA)

4.9 SUMMARY OF GRAPHICAL METHOD

Step 1: Convert the inequality constraint as equations and find co-ordinates of the line.

Step 2: Plot the lines on the graph.

(**Note:** If the constraint is \geq type, then the solution zone lies away from the centre. If the constraint is \leq type, then solution zone is towards the centre.)

Step 3: Obtain the feasible zone.

Step 4: Find the co-ordinates of the objectives function (profit line) and plot it on the graph representing it with a dotted line.

Step 5: Locate the solution point.

(**Note:** If the given problem is maximization, z_{\max} then locate the solution point at the far most point of the feasible zone from the origin and if minimization, Z_{\min} then locate the solution at the shortest point of the solution zone from the origin).

Step 6: Solution type

- i. If the solution point is a single point on the line, take the corresponding values of x_1 and x_2 .
- ii. If the solution point lies at the intersection of two equations, then solve for x_1 and x_2 using the two equations.
- iii. If the solution appears as a small line, then a multiple solution exists.
- iv. If the solution has no confined boundary, the solution is said to be an unbound solution.

Linear Programming: Graphical Method

Example 11: Solve the Geetha perfume company (Example 1.7) graphically using computer.

The formulated LP model is,

$$Z_{\max} = 7x_1 + 5x_2$$

Subject to constraints,

$$8x_1 + 4x_2 \leq 20 \quad \dots \dots \dots \text{(i)}$$

$$-x_1 + x_2 \leq 2 \quad \dots \dots \dots \text{(iii)}$$

where $x_1, x_2 \geq 0$

Solution: The input values of the problem are given to obtain the output screen as shown in Figure 4.7.

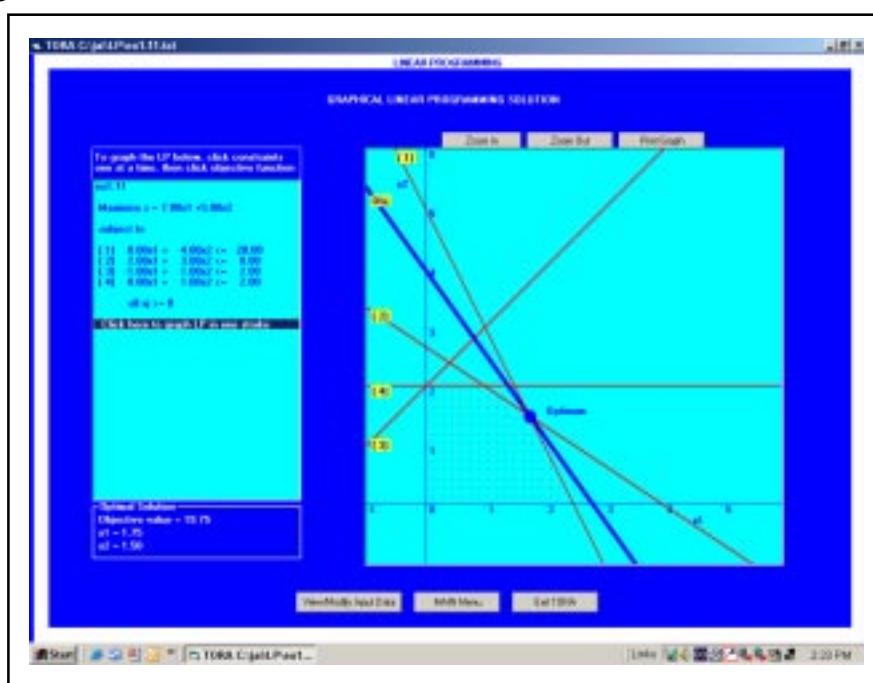


Figure 4.7: Graphical Presentation (Output Screen, TORA)

Results:

Perfumes to be produced, $x_1 = 1.75$ litres or 17.5 say 18 bottles of 100 ml each.

Body sprays to be produced, $x_1 = 1.50$ litres or 15 bottles of 100 ml each

Maximum profit, Z_{\max} = Rs. 19.75

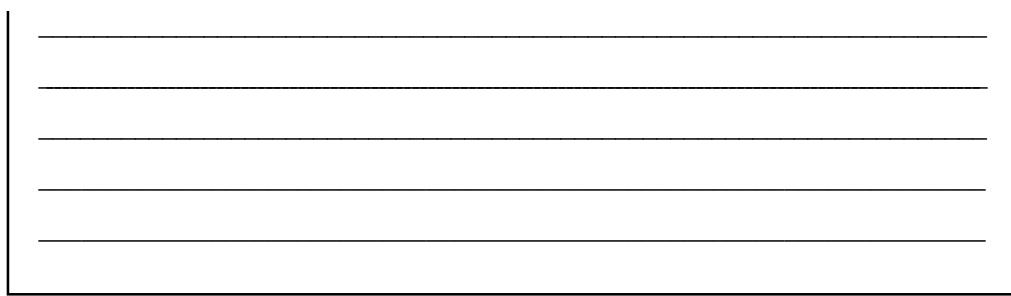
Check Your Progress 4.2

Discuss the limitations of graphical method in solving LPP.

- Notes:**

 - (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

Contd....



4.10 UNBOUNDED LP PROBLEM

Example 12: Solve the following LPP graphically

$$Z_{\max} = 6x_1 + 10x_2$$

Subject to constraints,

$x_2 \geq 10$ (ii)

$$2x_1 + 4x_2 \geq 20$$

where

Solution:

equity

$$x_1 = 6$$

$$x_2 = 10$$

1 2

$n = 6$ passes through $(6, 0)$

$x_1 = 10$ passes through $(0, 10)$

$$2x + 4x = 20 \text{ passes through } (10, 0), (5, 0)$$

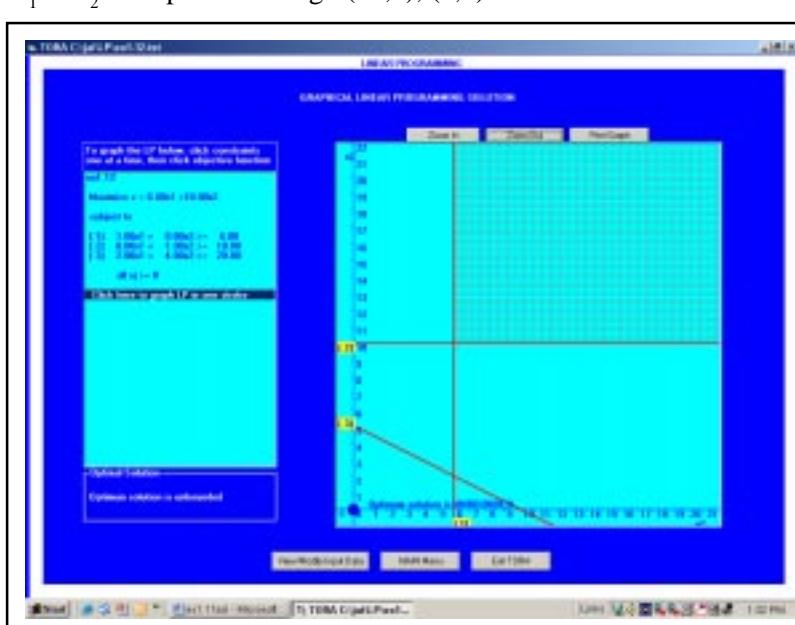


Figure 4.8: Graphical Presentation (Output Screen, TORA)

The given problem is maximization one. The solution point should be located at the furthermost point of the feasible region.

The feasible zone (shaded area) shown in Figure 4.8 is open-ended, i.e., it has no confined boundary. This means that the maximization is not possible or the LPP has no finite solution, and hence the solution is unbounded.

Linear Programming: Graphical Method

Example 13: Solve the given linear programming problem graphically using a computer.

$$\text{Maximize } Z = 3x_1 + 2x_2$$

Subject to constraints

$$x_1 + x_2 \geq 3 \quad \dots \dots \dots \text{(ii)}$$

$$x_1, x_2 \geq 0$$

Solution: The input as required is entered into the TORA input screen, the following output is obtained as shown in Figure 4.9 which shows that the solution is unbounded.

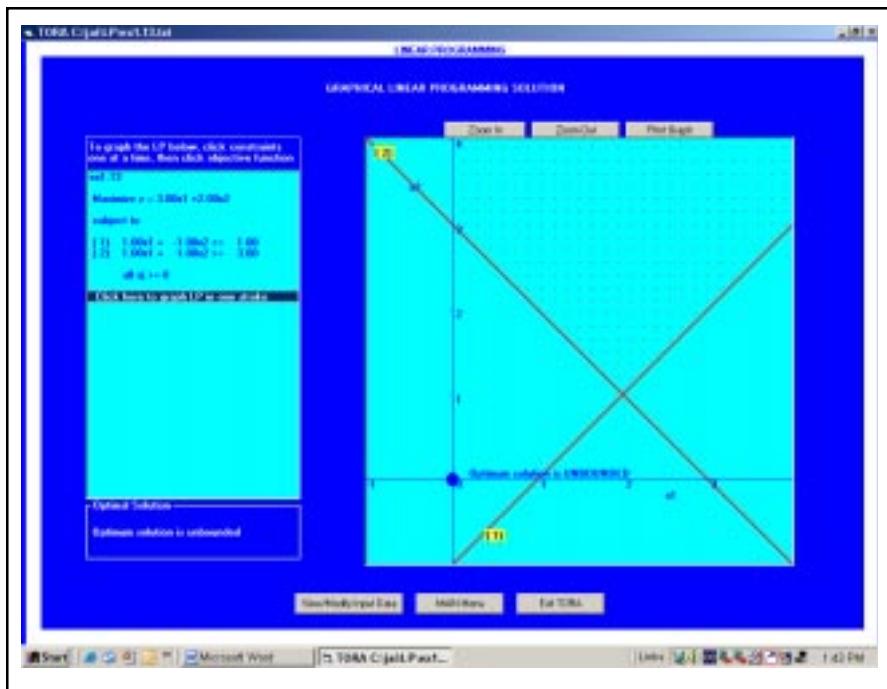


Figure 4.9: Graphical LP Solution (Output Screen, TORA)

Check Your Progress 4.3

What is unbound LP problem?

Notes: (a) Write your answer in the space given below.

- (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Chek Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

4.11 LET US SUM UP

Thus we can say that LP is a method of planning whereby objective function is maximised or minimised while at the same time satisfying the various restrictions placed on the potential solution. In technical words, linear programming is defined as a methodology whereby a linear function is optimized (minimised or maximised) subject to a set of linear constraints in the form of equalities or inequalities. Thus LP is a planning technique of selecting the best possible (optimal) strategy among number of alternatives.

4.12 LESSON-END ACTIVITY

LP is about trying to get the best result (e.g. maximum profit, least effort etc.) given some list of constraints Linear Programming allows for the ethical allocation of scarce or costly resources while still meeting all technical parameters. Explain how LP programmes are being used in self-diverse industries an sausage making, fruit juice mixing, baby cereals and milks, health foods, soups. Also facilitates in formulating recipes.

4.13 KEYWORDS

Linear Programming

Graphical Method

Maximisation

Minimisation

Constraints

Profit

Optimality

4.14 QUESTIONS FOR DISCUSSION

1. **Write True or False against each statement:**

- (a) LP is a widely used mathematical modeling technique.
- (b) LP consists of linear objectives and linear constraints.
- (c) Divisibility refers to the aim to optimize.
- (d) Limited resources means limited number of labour, material equipment and finance.
- (e) The objective function represents the aim or goal of the system, which has to be determined from the solution.

2. **Briefly comment on the following statements:**

- (a) Formulation of LP is the representation of problem situation in a mathematical form.
- (b) A model must have an objective function.
- (c) When feasible zone lies towards the origin.
- (d) LP techniques are used to optimize the resource for best result.
- (e) LP techniques are used in analyzing the effect of changes.

3. Fill in the blanks:

- Organization normally have _____ resources.
- A model has a _____ constraint.
- In real life, the two _____ problems are practiced very little.
- _____ refer to the products, workers', efficiency, and machines are assumed to be identical.
- The _____ function represents the aim or goal of the system.

4.15 TERMINAL QUESTIONS

- Define Linear Programming.
- What are the essential characteristics required for a linear programming model?
- What is meant by objective function in LP model?
- What is a constraint? Give a few examples of constraints in real life situations.
- Enumerate the steps involved in solving a LPP by graphical approach.
- What is the major limitation of the graphical method?
- List out the various constraint types in formulating a LP model.
- Define the feasible area.
- What are the possible solution types that can result in the graphical method?
- What is meant by an unbounded solution?
- How are multiple solutions interpreted in the graphical method?

Exercise Problems

- For the problem given in Example 7, formulate the constraints for the following without any change in R.H.S.:
 - The flower extract F_1 must be used at most to 15 litres and at least 5 litres.
 - The demand for perfume cannot be less than the demand for body spray.
 - The daily demand of body spray exceeds that of perfume by at least 2 litres.
- For the problem given in Example 1.7, determine the best feasible solution among the following values of x_1 and x_2 :
 - $x_1 = 2, x_2 = 1$
 - $x_1 = 0, x_2 = 3$
 - $x_1 = 3, x_2 = 1$
 - $x_1 = 5, x_2 = 1$
 - $x_1 = 2, x_2 = -1$
 - $x_1 = 1.75, x_2 = 1.50$
- Determine the feasible space for each of the following constraints:
 - $2x_1 - 2x_2 \leq 5$
 - $5x_1 + 10x_2 \leq 60$
 - $x_1 - x_2 \leq 0$
 - $4x_1 + 3x_2 \geq 15$

- (e) $x_2 \leq 5$
- (f) $x_1 \leq 30$
4. A company manufactures two types of products, A and B. Each product uses two processes, I and II. The processing time per unit of product A on process I is 6 hours and on the process II is 5 hours. The processing time per unit of product B on process I is 12 hours and on process II is 4 hours. The maximum number of hours available per week on process I and II are 75 and 55 hours respectively. The profit per unit of selling A and B are Rs.12 and Rs.10 respectively.
- Formulate a linear programming model so that the profit is maximized.
 - Solve the problem graphically and determine the optimum values of product A and B.
5. Formulate the following data as a linear programming model.

Products	Time required (minutes/unit)			Profit
	Lathe	Drilling	Cleaning	
A	25	30	15	25
B	15	5	10	30
C	20	15	10	50
Hours Available	250	400	200	

6. A nutrition scheme for babies is proposed by a committee of doctors. Babies can be given two types of food (I and II) which are available in standard sized packets, weighing 50 gms. The cost per packet of these foods are Rs. 2 and Rs. 3 respectively. The vitamin availability in each type of food per packet and the minimum vitamin requirement for each type of vitamin are summarized in the table given. Develop a linear programming model to determine the optimal combination of food type with the minimum cost such that the minimum requirement of vitamin is each type is satisfied.

Details of food type			
Vitamin availability per product			
Vitamin	Food Type I	Food Type II	Minimum Daily requirement
1	1	1	6
2	7	1	14
Cost/Packet (Rs.)	2	3	

7. Formulate the problem as a LP model

Resources/Constraints	Products/unit		Availability
	A	B	
Budget (Rs.)	8	4	4000
Machine Time	2	1	1000 hours
Assembly Time	3	4	750 hours
Selling Price	Rs. 20	Rs. 40	
Cost Price	Rs. 5	Rs. 20	

8. Solve the Chandru Bag company problem graphically.

- (a) Determine the values of x_1 , x_2 and Z_{\max} .

- (b) If the company has increased the demand for ordinary bag from 100 to 150, what is the new Z_{\max} value?
 - (c) If the demand for deluxe bags has reduced to 50 bags, determine the optimal profit value.
9. Solve the following linear programming model graphically:

$$\text{Maximize } Z = 30x_1 + 100x_2$$

Subject to constraints,

$$4x_1 + 6x_2 \leq 90$$

$$8x_1 + 6x_2 \leq 100$$

$$5x_1 + 4x_2 \leq 80$$

where $x_1, x_2 \geq 0$

10. Solve the following LP graphically:

$$\text{Maximize } Z = 8x_1 + 10x_2$$

Subject to constraints,

$$2x_1 + 3x_2 \geq 20$$

$$4x_1 + 2x_2 \geq 25$$

where $x_1, x_2 \geq 0$

11. Solve the two variable constraints using graphical method.

$$\text{Maximize } Z = 50x_1 + 40x_2$$

Subject to constraints

$$x_1 \geq 20$$

$$x_2 \leq 25$$

$$2x_1 + x_2 \leq 60$$

where $x_1, x_2 \geq 0$

12. Solve the following LP graphically using TORA.

$$\text{Maximize } Z = 1200x_1 + 1000x_2$$

Subject to constraints,

$$10x_1 + 4x_2 \geq 600$$

$$7x_1 + 10x_2 \geq 300$$

$$2x_1 + 4x_2 \leq 1000$$

$$9x_1 + 7x_2 \leq 2500$$

$$5x_1 + 4x_2 \leq 1200$$

where $x_1, x_2 \geq 0$

13. Solve graphically:

$$\text{Maximize } Z = 2x_1 + 3x_2$$

Subject to constraints,

$$x_1 - x_2 \leq 0$$

$$-3x_1 + x_2 \leq 25$$

where $x_1, x_2 \geq 0$

14. Solve the following LP graphically:

$$\text{Maximize } Z = 8x_1 + 10x_2$$

Subject to constraints,

$$0.5x_1 + 0.5x_2 \leq 150$$

$$0.6x_1 + 0.4x_2 \leq 145$$

$$x_1 \geq 30$$

$$x_1 \leq 150$$

$$x_2 \geq 40$$

$$x_2 \leq 200$$

where $x_1, x_2 \geq 0$

15. Determine the optimal values of x_1 and x_2 and hence find the maximum profits for the following LP problem:

$$\text{Maximize } Z = 4x_1 + 5x_2$$

Subject to constraints

$$x_1 + 3x_2 \leq 2$$

$$4x_1 + 5x_2 \leq 6$$

where $x_1, x_2 \geq 0$

4.16 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

1. (a) True (b) True (c) False (d) True (e) False
3. (a) scarce (b) non-negative (c) variable (d) Homogeneity
(e) objective

4.17 SUGGESTED READINGS

William H, *Model Building in Mathematical Programming*, Wiley Newyork.

Rohn E., "A New LP Approach to Bond Portfolio Management", Journal of Financial & Quantitative Analysis 22 (1987): 439-467.

Wagner H, *Principles of OR*, 2nd ed. Englewood Cliffs, N.J: Prentice Hall, 1975.

Moondra S., "An LP Model for Workforce Scheduling in Banks", Journal of Bank Research (1976).

LESSON

5

LINEAR PROGRAMMING: SIMPLEX METHOD

CONTENTS

- 5.0 Aims and Objectives
- 5.1 Introduction
- 5.2 Additional Variables used in Solving LPP
- 5.3 Maximization Case
- 5.4 Solving LP Problems Using Computer with TORA
- 5.5 Minimization LP Problems
- 5.6 Big M Method
- 5.7 Degeneracy in LP Problems
- 5.8 Unbounded Solutions in LPP
- 5.9 Multiple Solutions in LPP
- 5.10 Duality in LP Problems
- 5.11 Sensitivity Analysis
- 5.12 Let us Sum Up
- 5.13 Lesson-end Activities
- 5.14 Keywords
- 5.15 Questions for Discussion
- 5.16 Terminal Questions
- 5.17 Model Answers to Questions for Discussion
- 5.18 Suggested Readings

5.0 AIMS AND OBJECTIVES

In the previous lesson we have learnt linear programming with the help of graphical now we will learn the linear programming with the help of Simplex Method using minimization and maximization problems and the degeneracy in LP problems and also the Duality and Sensitivity Analysis.

5.1 INTRODUCTION

In practice, most problems contain more than two variables and are consequently too large to be tackled by conventional means. Therefore, an algebraic technique is used to solve large problems using Simplex Method. This method is carried out through iterative process systematically step by step, and finally the maximum or minimum values of the objective function are attained.

The basic concepts of simplex method are explained using the Example 1.8 of the packaging product mix problem illustrated in the previous chapter. The simplex method solves the linear programming problem in iterations to improve the value of the objective function. The simplex approach not only yields the optimal solution but also other valuable information to perform economic and 'what if' analysis.

5.2 ADDITIONAL VARIABLES USED IN SOLVING LPP

Three types of additional variables are used in simplex method such as,

- (a) Slack variables ($S_1, S_2, S_3, \dots, S_n$): Slack variables refer to the amount of unused resources like raw materials, labour and money.
 - (b) Surplus variables ($-S_1, -S_2, -S_3, \dots, -S_n$): Surplus variable is the amount of resources by which the left hand side of the equation exceeds the minimum limit.
 - (c) Artificial Variables ($a_1, a_2, a_3, \dots, a_n$): Artificial variables are temporary slack variables which are used for purposes of calculation, and are removed later.

The above variables are used to convert the inequalities into equality equations, as given in the Table 5.1 below.

Table 5.1: Types of Additional Variables

	Constraint Type	Variable added	Format
a)	Less than or equal to \leq	Add Slack Variable	+S
b)	Greater than or equal to \geq	Subtract surplus variable and add artificial variable	-S+a
c)	Equal to $=$	Add artificial variable	+a

5.3 MAXIMIZATION CASE

The packaging product mix problem is solved using simplex method.

$$\text{Maximize } Z = 6x_1 + 4x_2$$

Subject to constraints.

where $x_1, x_2 \geq 0$

Considering the constraint for cutting machine,

$$2x_1 + 3x_2 \geq 120$$

The inequality indicates that the left-hand side of the constraints equation has some amount of unused resources on cutting machine. To convert this inequality constraint into an equation, introduce a slack variable, S_3 which represents the unused resources. Introducing the slack variable, we have the equation

$$2x_1 + 3x_2 + S_3 = 120$$

Similarly for pinning machine, the equation is

$$2x_1 + x_2 + S_1 = 60$$

The variables S_3 and S_4 are known as slack variables corresponding to the three constraints. Now we have in all four variables (which includes slack variable) and two equations. If any two variables are equated to zero, we can solve the three equations of the system in two unknowns.

If variables x_1 and x_2 are equated to zero,

i.e., $x_1 = 0$ and $x_2 = 0$, then

$$S_3 = 120$$

$$S_4 = 60$$

This is the basic solution of the system, and variables S_3 and S_4 are known as Basic Variables, S_B while x_1 and x_2 known as Non-Basic Variables. If all the variables are non-negative, a basic feasible solution of a linear programming problem is called a Basic Feasible Solution.

Rewriting the constraints with slack variables gives us,

$$Z_{\max} = 6x_1 + 4x_2 + 0S_3 + 0S_4$$

Subject to constraints,

$$2x_1 + 3x_2 + S_3 = 120 \quad \dots \dots \dots \text{(i)}$$

$$2x_1 + x_2 + S_4 = 60 \quad \dots \dots \dots \text{(ii)}$$

where $x_1, x_2 \geq 0$

Though there are many forms of presenting Simplex Table for calculation, we represent the coefficients of variables in a tabular form as shown in Table 5.2.

Table 5.2: Co-efficients of Variables

Iteration Number	Basic Variables	Solution Value	X_1 K_C	X_2	S_3	S_4	Minimum Ratio	Equation
0	S_3	120	2	3	1	0	60	
	S_4	60	2	1	0	1	30	
	$- Z_j$	0	-6	-4	0	0		

If the objective of the given problem is a maximization one, enter the co-efficient of the objective function Z_j with opposite sign as shown in Table 5.3. Take the most negative coefficient of the objective function and that is the key column K_c . In this case, it is -6. Find the ratio between the solution value and the key column coefficient and enter it in the minimum ratio column. The intersecting coefficients of the key column and key row are called the pivotal element i.e. 2. The variable corresponding to the key column is the entering element of the next iteration table and the corresponding variable of the key row is the leaving element of the next iteration table. In other words, x_1 replaces S_4 in the next iteration table. Table 5.3 indicates the key column, key row and the pivotal element.

Table 5.3

Iteration Number	Basic Variables	Solution Value	X_1 K_C	X_2	S_3	S_4	Minimum Ratio	Equation
0	S_3	120	2	3	1	0	60	
K_r	S_4	60	2	1	0	1	30	
	$-Z_j$	0	-6	-4	0	0		

In the next iteration, enter the basic variables by eliminating the leaving variable (i.e., key row) and introducing the entering variable (i.e., key column). Make the pivotal element as 1 and enter the values of other elements in that row accordingly. In this case, convert the pivotal element value 2 as 1 in the next iteration table. For this, divide the pivotal element by 2. Similarly divide the other elements in that row by 2. The equation is $S_4 / 2$. This row is called as Pivotal Equation Row Pe. The other co-efficients of the key column in iteration Table 5.4 must be made as zero in the iteration Table 5.5. For this, a solver, Q, is formed for easy calculation. Change the sign of the key column coefficient, multiply with pivotal equation element and add with the corresponding variable to get the equation,

$$\text{Solver, } Q = S_B + (-K_c \times P_e)$$

The equations for the variables in the iteration number 1 of table 8 are,

$$\begin{aligned} \text{For } S_3, Q &= S_B + (-K_c \times P_e) \\ &= S_3 + (-2x P_e) \\ &= S_3 - 2P_e \end{aligned} \quad \dots \dots \dots \quad (1)$$

$$\begin{aligned}
 \text{For } -Z, \quad Q &= S_B + (-K_c \times P_e) \\
 &= -Z + ((-6) \times P_e) \\
 &= -Z + 6P_e \quad \dots \dots \dots \text{(iii)}
 \end{aligned}$$

Using the equations (i) and (ii) the values of S_3 and $-Z$ for the values of Table 1 are found as shown in Table 5.4

Table 5.4: S₃ and -Z Values Calculated

Iteration Number	Basic Variables	Solution Value	X_1 K_C	X_2 K_C	S_3	S_4	Minimum Ratio	Equation
0	S_3	120	2	3	1	0	60	
K_r	S_4	60	2	1	0	1	30	
	$-Z_j$	0	-6	-4	0	0		
I K_r	S_3	60	0	2	1	-1	30	$S_3 - 2P_e$
P_e	x_1	30	1	$\frac{1}{2}$	0	$\frac{1}{2}$	60	$S_{4/2}$
	$-Z_j$	100	0	-1	0	3		$-Z + 6P_e$

Using these equations, enter the values of basic variables S_B and objective function Z. If all the values in the objective function are non-negative, the solution is optimal. Here, we have one negative value – 1. Repeat the steps to find the key row and pivotal equation values for the iteration 2 and check for optimality.

In the iteration 2 number of Table 5.5, all the values of Z_j are non-negative, $Z_j \geq 0$, hence optimality is reached. The corresponding values of x_1 and x_2 for the final iteration table gives the optimal values of the decision variables i.e., $x_1 = 15$, $x_2 = 30$. Substituting these values in the objectives function equation, we get

$$\begin{aligned}
 Z_{\max} &= 6x_1 + 4x_2 \\
 &= 6(15) + 4(30) \\
 &= 90 + 120 \\
 &\equiv \text{Rs. } 210,00
 \end{aligned}$$

Table 5.5: Iteration Table

Iteration Number	Basic Variables	Solution Value	X_1	X_2	S_3	S_4	Minimum Ratio	Equation
0 K_r	S_3	120	2	3	1	0	60	
	S_4	60	2	1	0	1	30	
	$-Z_j$	0	-6	-4	0	0		
1 K_r P_e	S_3	60	0	2	1	-1	30	$S_3 - 2P_e$
	x_1	30	1	$\frac{1}{2}$	0	$\frac{1}{2}$	60	$S_{4/2}$
	$-Z_j$	100	0	-1	0	3		$-Z + 6P_e$
2 P_e	X_2	30	0	1	$\frac{1}{2}$	-		$S_{3/2}$
	x_1	15	1	0	-	1/2		$S_3 - P_{e/2}$
	$-Z_j$	210	0	0	1/4	$\frac{3}{4}$	5/2	$-Z + P_e$

The solution is,

$x_1 = 15$ corrugated boxes are to be produced and

$x_2 = 30$ carton boxes are to be produced to yield a

Profit, $Z_{\max} = \text{Rs. } 210.00$

Summary of LPP Procedure

Step 1: Formulate the LP problem.

Step 2: Introduce slack /auxiliary variables.

if constraint type is \leq introduce $+ S$

if constraint type is \geq introduce $- S + a$ and

if constraint type is $=$ introduce a

Step 3: Find the initial basic solution.

Step 4: Establish a simplex table and enter all variable coefficients. If the objective function is maximization, enter the opposite sign co-efficient and if minimization, enter without changing the sign.

Step 5: Take the most negative coefficient in the objective function, Z_j to identify the key column (the corresponding variable is the entering variable of the next iteration table).

Step 6: Find the ratio between the solution value and the coefficient of the key column. Enter the values in the minimum ratio column.

Step 7: Take the minimum positive value available in the minimum ratio column to identify the key row. (The corresponding variable is the leaving variable of the table).

Step 8: The intersection element of the key column and key row is the pivotal element.

Step 9: Construct the next iteration table by eliminating the leaving variable and introducing the entering variable.

Step 10: Convert the pivotal element as 1 in the next iteration table and compute the other elements in that row accordingly. This is the pivotal equation row (not key row).

Step 11: Other elements in the key column must be made zero. For simplicity, form the equations as follows: Change the sign of the key column element, multiply with pivotal equation element and add the corresponding variable.

Step 12: Check the values of objective function. If there are negative values, the solution is not an optimal one; go to step 5. Else, if all the values are positive, optimality is reached. Non-negativity for objective function value is not considered. Write down the values of x_1, x_2, \dots, x_i and calculate the objective function for maximization or minimization.

Note:

- (i) If there are no x_1, x_2 variables in the final iteration table, the values of x_1 and x_2 are zero.
- (ii) Neglect the sign for objective function value in the final iteration table.

5.4 SOLVING LP PROBLEMS USING COMPUTER WITH TORA

From the MAIN MENU, select LINEAR PROGRAMMING option, and enter the input values of the previously discussed problem as shown in the Figure 5.1.

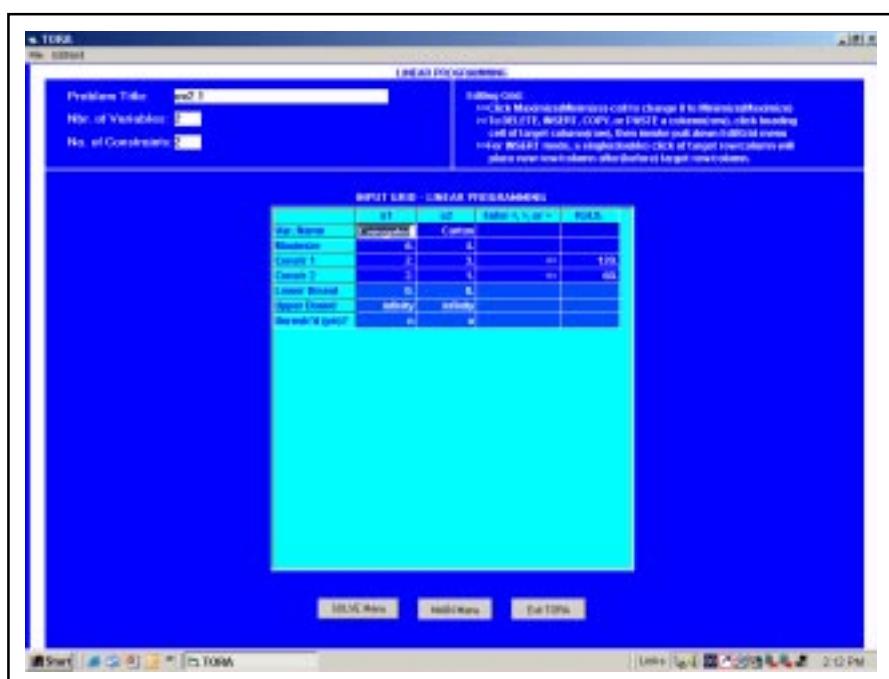


Figure 5.1: Solving LPP using Computer with TORA (Input Screen)

Click Solve Menu, and select Solve Problem → Algebraic → Iterations → All-Slack Starting Solution. Now, click Go To Output screen, then the first iteration table will be displayed. To select the entering variable, click a non-basic variable (if correct, the column turns green). Similarly, select the leaving variable (if correct, the row turns red), Figure 5.2.

Iteration 1						Conjugated	Carton				Solution
Basic		x1	x2	sx3	sx4						
z (max)		-6.00	-4.00	0.00	0.00						0.00
sx3		2.00	3.00	1.00	0.00						120.00
sx4		2.00	1.00	0.00	1.00						60.00
Lower Bound		0.00	0.00								
Upper Bound		infinity	infinity								
Unrest'r'd [y/n]?		n	n								

Figure 5.2: Selecting the Leaving Variable (TORA, Output Screen)

Then click Next Iteration button to display the next iteration table as shown in Figure 5.3.

Next Iteration							All Iterations	Write to Printer
Iteration 1	Corrugated	Carton	x1	x2	sx3	sx4	Solution	
Basic			-6.00	-4.00	0.00	0.00	0.00	
z (max)								
sx3			2.00	3.00	1.00	0.00	120.00	
sx4			2.00	1.00	0.00	1.00	60.00	
Lower Bound			0.00	0.00				
Upper Bound			infinity	infinity				
Unrestrl'd (y/n)?			n	n				
Iteration 2	Corrugated	Carton	x1	x2	sx3	sx4	Solution	
Basic			0.00	1.00	0.00	3.00	180.00	
z (max)								
sx3			0.00	2.00	1.00	-1.00	60.00	
x1			1.00	0.50	0.00	0.50	30.00	
Lower Bound			0.00	0.00				
Upper Bound			infinity	infinity				
Unrestrl'd (y/n)?			n	n				

Figure 5.3: Next Iteration Table (TORA, Output Screen)

Again click next iteration button to get the third and final iteration table. A pop-up menu also indicates that the solution has reached the optimal level. Now we can notice that all the values in the objective function Z_{\max} row are non-negative which indicates that the solution is optimal. The final Iteration Table is shown in Figure 5.4.

The screenshot shows the TORA software interface with the title 'LINEAR PROGRAMMING'. In the center, there is a table titled 'Iteration 3' with columns for Corrugated, Carton, x1, x2, sx3, sx4, and Solution. The 'Solution' column shows values: 0.00, 1.00, 0.50, 30.00, 0.00, 0.00, and 210.00 respectively. Below the table, a small window titled 'Optimal Solution' displays the message 'Iteration 3 is optimal' with an 'OK' button. At the bottom of the screen, there is a taskbar with icons for Start, Microsoft Word, and Microsoft Excel, along with other system status indicators.

Iteration 3						
	Corrugated	Carton	x1	x2	sx3	sx4
Basic			0.00	1.00	0.50	30.00
x1			0.00	0.00	0.00	0.00
x2			0.00	0.00	0.00	0.00
Lower Bound			infinity	infinity		
Upper Bound			infinity	infinity		
Unrestrl'd (y/n)?			n	n		

Figure 5.4: Final Iteration Table (TORA, Output Screen)

From the final Iteration Table, the values of X_1 , X_2 and Z_{\max} are taken to the corresponding values in the solution column (last column) of the simplex table.

$$\text{i.e., } Z_{\max} = 210.00$$

$$X_1 = 30.00$$

$$X_2 = 15.00$$

Example 1: Solve the LP problem using Simplex method. Determine the following :

- (a) What is the optimal solution?
- (b) What is the value of the objective function?

- (c) Which constraint has excess resources and how much?

$$Z_{\max} = 5x_1 + 6x_2$$

Subject to constraints,

$$2x_1 + x_2 \leq 2000 \quad \dots \dots \dots \text{(i)}$$

$$x_1 \leq 800 \quad \dots \dots \dots \text{(ii)}$$

$$x_2 \leq 200 \quad \dots \dots \dots \text{(iii)}$$

$$\text{where } x_1, x_2 \geq 0$$

Solution: Converting the inequality constraints by introducing the slack variables,

$$Z_{\max} = 5x_1 + 6x_2 + 0S_3 + 0S_4 + 0S_5$$

$$2x_1 + x_2 + S_3 = 2000$$

$$x_1 + S_4 = 800$$

$$x_2 + S_5 = 200$$

Equate x_1 and x_2 to zero , to find the initial basic solution

$$2(0) + 0 + S_3 = 2000$$

$$0 + S_4 = 800$$

$$0 + S_5 = 200$$

The initial basic solution is,

$$S_3 = 2000$$

$$S_4 = 800$$

$$S_5 = 200$$

Establish a simplex table to represent the co-efficient of variables for optimal computation as shown in Table 5.6.

Table 5.6: Simplex Table

Iteration Number	Basic Variable	Solution Value	X ₁	X ₂	S ₃	S ₄	S ₅	Min Ratio	Equation
0	S ₃	2000	2	1	1	0	0	2000	
	S ₄	800	1	0	0	1	0	∞	
K _r	S ₃	200	0	1	0	0	1	200	
	-Z	1200	-5	-6	0	0	0		
1	S ₃	1800	2	0	1	0	-1	900	S ₃ - P _e
K _r	S ₄	800	1	0	0	1	0	800	S ₄
P _e	X ₂	200	0	1	0	0	1	∞	S ₅
	-Z	1200	-5	0	0	0	6		-Z + 6P _e
2	S ₃	200	0	-2	1	-2	-1		S ₃ - 2P _e
P _e	X ₁	800	1	0	0	1	0		S ₄
	X ₂	200	0	1	0	0	1		X ₂
	-Z _j	5200	0	0	0	5	6		-Z + 5P _e

In the final table, all the values of $-Z_j$ are ≥ 0 , hence optimality is reached. The optimum solution is,

- (a) The value of $x_1 = 800$ units

$$x_2 = 200 \text{ units}$$

- $$\begin{aligned}
 \text{(b) Objective function } Z_{\max} &= 5x_1 + 6x_2 \\
 &= 5(800) + 6(200) \\
 &\equiv \text{Rs. 5200.00}
 \end{aligned}$$

- (c) In the final iteration Table 5.2, slack variable S_3 represents the first constraint, therefore this constraint has excess unused resources of 200 units.

5.5 MINIMIZATION LP PROBLEMS

In real life we need to minimize cost or time in certain situations. The objective now is minimization. Procedure for minimization problems is similar to maximization problems. The only difference is, enter the coefficients of the objective function in the simplex table without changing the sign.

Another way to solve minimization problems is by converting the objective function as a maximization problem by multiplying the equation by (-1) .

For example, if the objective function is,

$$\text{Minimize } Z = 10x_1 + 5x_2$$

Convert the objective function into maximization and solve

$$\text{Maximize } Z = -10x_1 - 5x_2$$

5.6 BIG M METHOD

So far, we have seen the linear programming constraints with less than type. We come across problems with ‘greater than’ and ‘equal to’ type also. Each of these types must be converted as equations. In case of ‘greater than’ type, the constraints are rewritten with a negative surplus variable S_1 and by adding an artificial variable a . Artificial variables are simply used for finding the initial basic solutions and are thereafter eliminated. In case of an ‘equal to’ constraint, just add the artificial variable to the constraint.

The co-efficient of artificial variables a_1, a_2, \dots are represented by a very high value M , and hence the method is known as BIG-M Method.

Example 2: Solve the following LPP using Big M Method.

$$\text{Minimize } Z = 3x_1 + x_2$$

Subject to constraints

$$5x_1 + 3x_2 \geq 7 \quad \dots \dots \dots \text{(ii)}$$

$$3x_1 + 2x_2 \leq 6 \quad \dots \dots \dots \text{(iii)}$$

where $x_1, x_2 \geq 0$

Solution: Introduce slack and auxiliary variables to represent in the standard form. Constraint $4x_1 + x_2 = 4$ is introduced by adding an artificial variable a_1 , i.e.,

$$4x_1 + x_2 + a_1 = 4$$

Constraint, $5x_1 + 3x_2 \geq 7$ is converted by subtracting a slack S_1 and adding an auxiliary variable a_1 .

$$5x_1 + 3x_2 - S_1 + a_2 = 7$$

Constraint $3x_1 + 2x_2 \leq 6$ is included with a slack variable S_2 .

$$3x_2 + 2x_3 + S_3 = 6$$

The objective must also be altered if auxiliary variables exist. If the objective function is minimization, the co-efficient of auxiliary variable is +M (and -M, in case of maximization)

The objective function is minimization,

$$\text{Minimize } Z = 3x_1 + x_2 + 0S_1 + 0S_2 + Ma_1 + Ma_2$$

$$Z_{\min} = 3x_1 + x_2 + Ma_1 + Ma_2$$

The initial feasible solution is (Put $x_1, x_2, S_1 = 0$)

$$a_1 = 4$$

$$a_2 = 7$$

$$s_2 = 6$$

Establish a table as shown below and solve:

Table 5.7: Simplex Table

Iteration Number	Basic Variables	Solution Value	X_1	X_2	S_1	S_2	a_1	a_2	Min Ratio	Equation
0 Kr	Z	0	3	1	0	0	M	M	0.75	$Z + (-Ma_1) - Ma_2$
	a_1	4	4	1	0	0	1	0		
	a_2	7	5	3	-1	0	0	1		
	S_2	6	3	2	0	1	0	0	2	
	Z^1	$-11M$	$9M + 3$	$-4M + 1$	M	0	0	0		
1 Pe Kr	X_1	1	1	$\frac{1}{4}$	0	0	4	$a_1/4$ $a_2 - 5Pe$ $S_2 - 3Pe$ $Z^1 + (9M - 3) Pe$	1.14	$a_1/4$ $a_2 - 5Pe$ $S_2 - 3Pe$ $Z^1 + (9M - 3) Pe$
	a_2	2	0	$\frac{7}{4}$	-1	0				
	S_2	3	0	$\frac{5}{4}$	0	1				
	Z^1	$-2M - 3$	0	$\frac{7M}{4} + \frac{1}{4}$	M	0				
2	x_1	$\frac{5}{7}$	1	0	$-\frac{1}{7}$	0	$\frac{a_2}{7/4}$	$X_1 - Pe/4$ $\frac{a_2}{7/4}$ $Z^1 + (7M/4 - 1/4) Pe$	$\frac{a_2}{7/4}$	$X_1 - Pe/4$ $\frac{a_2}{7/4}$ $Z^1 + (7M/4 - 1/4) Pe$
	x_2	$\frac{8}{7}$	0	1	$-\frac{4}{7}$	0				
	S_2	$\frac{22}{14}$	0	0	$\frac{10}{14}$	1				
	Z^1	$-\frac{23}{7}$	0	0	$\frac{1}{7}$	0				

The solution is,

$$x_1 = 5/7 \text{ or } 0.71$$

$$x_2 = 8/7 \text{ or } 1.14$$

$$\begin{aligned} Z_{\min} &= 3 \times 5/7 + 8/7 \\ &= 23/7 \text{ or } 3.29 \end{aligned}$$

Check Your Progress 5.1

- What are the different types of additional variables used in simplex method?
- How will you introduce/auxiliary variables in solving LPT problem?

Notes:

- (a) Write your answer in the space given below.
- (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
- (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

Linear Programming : Simplex Method

5.7 DEGENERACY IN LP PROBLEMS

In solving linear programming problem, while improving the basic solution, it may so transpire that there is no scope to generate an optimal solution. This is known as 'degeneracy' in linear programming. This occurs when there is a tie in the minimum ratio column. In other words, two or more values in the minimum ratio column are the same. To resolve degeneracy, the following method is used. Divide the key column values (of the tied rows) by the corresponding values of columns on the right side. This makes the values unequal and the row with minimum ratio is the key row.

Example 3: Consider the following LPP,

$$\text{Maximize } Z = 2x_1 + x_2$$

Subject to constraints,

$$4x_1 + 3x_2 \leq 12 \quad \dots \dots \dots \text{(i)}$$

$$4x_1 + x_2 \leq 8 \quad \dots \dots \dots \text{(ii)}$$

Solution: Converting the inequality constraints by introducing the slack variables.

Maximize $Z = 2x_1 + x_2$

Subject to constraints

$$4x_1 + 3x_2 + 5x_3 \equiv 12 \quad \dots \dots \dots \text{(iv)}$$

Equating $x_1, x_2 = 0$, we get

S = 12

5

4

Table 5.8: Illustrating Degeneracy

Iteration Number	Basic Variables	Solution Value	X ₁	X ₂	S ₃	S ₄	S ₅	Minimum Ratio	Equation
0	S ₃	12	4	3	1	0	0	3	
	S ₄	8	4	1	0	1	0	2	
	S ₅	8	4	-1	0	0	1	2	
	-Z	0	-2	-1	0	0	0		

For S_4 ; $4/2 = 2$ tie
For S_5 ; $4/2 = 2$

After entering all the values in the first iteration table, the key column is -2, variable corresponding is x_1 . To identify the key row there is tie between row S_4 and row S_5 with same values of 2, which means degeneracy in solution. To break or to resolve this, consider the column right side and divide the values of the key column values. We shall consider column x_2 , the values corresponding to the tie values 1, -1. Divide the key column values with these values, i.e., $1/4, -1/4$ which is 0.25 and -0.25. Now in selecting the key row, always the minimum positive value is chosen i.e., row S_4 . Now, S_4 is the leaving variable and x_1 is an entering variable of the next iteration table. The problem is solved. Using computer and the solution is given in the Figure 5.5.

The screenshot shows the TORA software interface with three iterations of the Simplex method:

- Iteration 1:**

	x_1	x_2	sx_3	sx_4	sx_5	Solution
z (max)	-2.00	-1.00	0.00	0.00	0.00	0.00
sx_3	4.00	3.00	1.00	0.00	0.00	12.00
sx_4	4.00	1.00	0.00	1.00	0.00	8.00
sx_5	4.00	-1.00	0.00	0.00	1.00	8.00
Lower Bound	0.00	0.00				
Upper Bound	infinity	infinity				
Unrestr'd (y/n)?	n	n				
- Iteration 2:**

	x_1	x_2	sx_3	sx_4	sx_5	Solution
z (max)	0.00	-0.50	0.00	0.50	0.00	4.00
sx_3	0.00	2.00	1.00	-1.00	0.00	4.00
x_1	1.00	0.25	0.00	0.25	0.00	2.00
sx_5	0.00	-2.00	0.00	-1.00	1.00	0.00
Lower Bound	0.00	0.00				
Upper Bound	infinity	infinity				
Unrestr'd (y/n)?	n	n				
- Iteration 3:**

	x_1	x_2	sx_3	sx_4	sx_5	Solution
z (max)	0.00	0.00	0.25	0.25	0.00	5.00
x_2	0.00	1.00	0.50	-0.50	0.00	2.00
x_1	1.00	0.00	-0.13	0.38	0.00	1.50
sx_5	0.00	0.00	1.00	-2.00	1.00	4.00
Lower Bound	0.00	0.00				
Upper Bound	infinity	infinity				
Unrestr'd (y/n)?	n	n				

Buttons at the bottom: View/Modify Input Data, MAIN Menu, Exit TORA.

Figure 5.5: LPP Solved Using Computer with TORA (Output Screen)

5.8 UNBOUNDED SOLUTIONS IN LPP

In a linear programming problem, when a situation exists that the value objective function can be increased infinitely, the problem is said to have an 'unbounded' solution. This can be identified when all the values of key column are negative and hence minimum ratio values cannot be found.

Table 5.9: Illustrating Unbounded Solution

Iteration Value	Basic Variable	Solution Value	X_1	X_2	S_3	S_4	S_5	Minimum Ratio	Equation
1	S_3	12	1	-2	1	0	0		
	X_1	8	3	-1	0	1	0		
	S_4	4	2	-4	0	0	1		
	-Z	0	-4	-8	0	0	0		

For S_3 ; $12/-2$
For X_1 ; $8/-1$ all values are negative
For S_4 ; $4/-4$

5.9 MULTIPLE SOLUTIONS IN LPP

In the optimal iteration table if $(P_j - Z_j)$ value of one or more non-basic variable is equal to 0, then the problem is said to have multiple or alternative solutions.

Table 5.10: Illustrating Multiple Solutions

P _j			4	1	0	0	0		
Iteration Number	Basic Number	Solution Value	X ₁	X ₂	S ₃	S ₄	S ₅	Minimum Ratio	Equation
2	X ₂	6	5	2	0	1	0		
	S ₂	3	4	1	2	1	0		
	Z _j	4	4	1	1	2	0		
P _j - Z _j			0	0	-1	-2	0		

5.10 DUALITY IN LP PROBLEMS

All linear programming problems have another problem associated with them, which is known as its dual. In other words, every minimization problem is associated with a maximization problem and vice-versa. The original linear programming problem is known as primal problem, and the derived problem is known as its dual problem. The optimal solutions for the primal and dual problems are equivalent.

Conversion of primal to dual is done because of many reasons. The dual form of the problem, in many cases, is simple and can be solved with ease. Moreover, the variables of the dual problem contain information useful to management for analysis.

Procedure

Step 1: Convert the objective function if maximization in the primal into minimization in the dual and vice versa. Write the equation considering the transpose of RHS of the constraints

Step 2: The number of variables in the primal will be the number of constraints in the dual and vice versa.

Step 3: The co-efficient in the objective function of the primal will be the RHS constraints in the dual and vice versa.

Step 4: In forming the constraints for the dual, consider the transpose of the body matrix of the primal problems.

Note: Constraint inequality signs are reversed.

Example 4: Construct the dual to the primal problem

$$\text{Maximize } Z = 6x_1 + 10x_2$$

Subject to constraints.

$$2x_1 + 8x_2 \leq 60 \quad \dots \dots \dots \text{(i)}$$

$$5x_s - 6x_u \leq 10 \quad \dots \dots \dots \text{(iii)}$$

$x_2 \leq 40$ (iv)

where $x_1, x_2 \geq 0$

Solution:

$$\text{Minimize } W = 60y_1 + 45y_2 + 10y_3 + 40y_4$$

Subject to constraints,

$$2y_1 + 3y_2 + 5y_3 + 0y_4 \geq 6$$

$$8y_1 + 5y_2 + 6y_3 + y_4 \geq 10$$

where $y_1, y_2, y_3, y_4 \geq 0$

Example 5: Construct a dual for the following primal

$$\text{Minimize } Z = 6x_1 - 4x_2 + 4x_3$$

Subject to constraints,

$$6x_1 - 10x_2 + 4x_3 \geq 14 \quad \dots \dots \dots \text{(i)}$$

$$6x_1 + 2x_2 + 6x_3 \geq 10 \quad \dots \dots \dots \text{(ii)}$$

$$7x_1 - 2x_2 + 5x_3 \leq 20 \quad \dots \dots \dots \text{(iii)}$$

where $x_1, x_2, x_3 \geq 0$

Solution: Convert 'less than' constraints into 'greater than' type by multiplying by

(-1) on both sides (i.e., for e.g. iii).

$$6x_1 - 10x_2 + 4x_3 \geq 14$$

$$6x_1 + 2x_2 + 6x_3 \geq 10$$

$$-7x_1 + 2x_2 - 5x_3 \geq 20$$

$$x_1 - 4x_2 + 5x_3 \geq 3$$

$$4x_1 + 7x_2 - 4x_3 \geq 20$$

The dual for the primal problem is,

$$\text{Maximize } W = 14y_1 + 10y_2 + 20y_3 + 3y_4 + 20y_5$$

Subject to constraints,

$$6y_1 + 6y_2 - 7y_3 + y_4 + 4y_5 \leq 6$$

$$10y_1 + 2y_2 + 2y_3 - 4y_4 + 7y_5 \leq -4$$

$$4y_1 + 6y_2 - 5y_3 + 5y_4 - 4y_5 \leq 4$$

where y_1, y_2, y_3, y_4 and $y_5 \geq 0$

5.11 SENSITIVITY ANALYSIS

Sensitivity analysis involves 'what if?' questions. In the real world, the situation is constantly changing like change in raw material prices, decrease in machinery availability, increase in profit on one product, and so on. It is important to decision makers for find out how these changes affect the optimal solution. Sensitivity analysis can be used to provide information and to determine solution with these changes.

Sensitivity analysis deals with making individual changes in the co-efficient of the objective function and the right hand sides of the constraints. It is the study of how changes in the co-efficient of a linear programming problem affect the optimal solution.

We can answer questions such as.

- i. How will a change in an objective function co-efficient affect the optimal solution?
 - ii. How will a change in a right-hand side value for a constraint affect the optimal solution?

Linear Programming : Simplex Method

For example, a company produces two products x_1 and x_2 with the use of three different materials 1, 2 and 3. The availability of materials 1, 2 and 3 are 175, 50 and 150 respectively. The profit for selling per unit of product x_1 is Rs. 40 and that of x_2 is Rs. 30. The raw material requirements for the products are shown by equations, as given below.

$$Z_{\max} = 40x_1 + 30x_2$$

Subject to constraints

$$4x_1 + 5x_2 \leq 175 \quad \dots \dots \dots \text{(i)}$$

$$6x_1 + 3x_2 \leq 150 \quad \dots \dots \dots \text{(iii)}$$

where $x_1, x_2 \geq 0$

The optimal solution is

x₁ = Rs. 12.50

x₂ = Rs. 25.00

$$Z = 40 \times 12.50 + 30 \times 25.00$$

- Rs 1250.00

The problem is solved using TORA software and the output screen showing sensitivity analysis is given in Table 5.11.

Change in objective function co-efficients and effect on optimal solution

Table 5.11: Sensitivity Analysis Table Output

The screenshot shows the TORA software interface with the title "LINEAR PROGRAMMING". The main window displays the following data:

Variable	Value	Obj. Cost	(M) Obj. Cost
x1	12.00	60.00	508.00
x2	36.00	36.00	254.00
Unused	19.0	Start. Supplies	
S1	175.00	0.00	
S2	50.00	0.00	
S3	150.00	0.00	

Below this, a "Sensitivity Analysis" table is shown:

Variable	Current Obj. Cost	Min Obj. Cost	Max Obj. Cost	Reduced Cost
x1	60.00	24.00	68.00	0.00
x2	36.00	28.00	58.00	0.00
Unused	19.0	Min FRS	Max FRS	Unit Price
S1	175.00	156.00	175.00	1.33
S2	50.00	50.00	Infinity	0.00
S3	150.00	150.00	262.50	4.44

At the bottom of the window, there are buttons for "View/Modify Input Data", "Help/Menu", "Edit TORA", and "Microphone". The taskbar at the bottom of the screen shows the TORA application is running.

Referring to the current objective co-efficient, if the values of the objective function co-efficient decrease by 16 (Min. obj. co-efficient) and increase by 20 (Max. obj. co-efficient) there will not be any change in the optimal values of $x_1 = 12.50$ and $x_2 = 25.00$. But there will be a change in the optimal solution, i.e. Z_{\max} .

Note: This applies only when there is a change in any one of the co-efficients of variables i.e., x_1 or x_2 . Simultaneous changes in values of the co-efficients need to apply for 100 Percent Rule for objective function co-efficients.

$$\begin{aligned} \text{For } x_1, \text{ Allowable decrease} &= \text{Current value} - \text{Min. Obj. co-efficient} \\ &= 40 - 24 \\ &= \text{Rs. 16} \end{aligned} \quad \text{----- (i)}$$

$$\begin{aligned} \text{Allowable increase} &= \text{Max. Obj. co-efficient} - \text{Current value} \\ &= 60 - 40 \\ &= \text{Rs. 20.00} \end{aligned} \quad \text{----- (ii)}$$

$$\begin{aligned} \text{Similarly, For } x_2, \text{ Allowable decrease} &= \text{Rs. 10.00} \\ \text{Allowable increase} &= \text{Rs. 20.00} \end{aligned} \quad \text{----- (iii) (iv)}$$

For example, if co-efficient of x_1 is increased to 48, then the increase is $48 - 40 = \text{Rs. 8.00}$. From (ii), the allowable increase is 20, thus the increase in x_1 co-efficient is $8/20 = 0.40$ or 40%.

Similarly,

If co-efficient of x_2 is decreased to 27, then the decrease is $30 - 27 = \text{Rs. 3.00}$.

From (iii), the allowable decrease is 10, thus the decrease in x_2 co-efficient is $3/10 = 0.30$ or 30%.

Therefore, the percentage of increase in x_1 and the percentage of decrease in x_2 is 40 and 30 respectively.

i.e. $40\% + 30\% = 70\%$

For all the objective function co-efficients that are changed, sum the percentage of the allowable increase and allowable decrease. If the sum of the percentages is less than or equal to 100%, the optimal solution does not change, i.e., x_1 and x_2 values will not change.

But Z_{\max} will change, i.e.,

$$= 48(12.50) + 27(25)$$

$$= \text{Rs. 1275.00}$$

If the sum of the percentages of increase and decrease is greater than 100%, a different optimal solution exists. A revised problem must be solved in order to determine the new optimal values.

Change in the right-hand side constraints values and effect on optimal solution

Suppose an additional 40 kgs of material 3 is available, the right-hand side constraint increases from 150 to 190 kgs.

Now, if the problem is solved, we get the optimal values as

$$x_1 = 23.61, x_2 = 16.11 \text{ and } Z_{\max} = 1427.78$$

From this, we can infer that an additional resources of 40 kgs increases the profit by $= 1427.78 - 1250 = \text{Rs. 177.78}$

Therefore, for one kg or one unit increase, the profit will increase by

$$= 177.78 / 40$$

$$= \text{Rs. 4.44}$$

Dual price is the improvement in the value of the optimal solution per unit increase in the right-hand side of a constraint. Hence, the dual price of material 3 is Rs 4.44 per kg.

Linear Programming : Simplex Method

Increase in material 2 will simply increase the unused material 2 rather than increase in objective function. We cannot increase the RHS constraint values or the resources. If the limit increases, there will be a change in the optimal values.

The limit values are given in Table 2.10, i.e., Min RHS and Max RHS values.

For example, for material 3, the dual price Rs. 4.44 applies only to the limit range 150 kgs to 262.50 kgs.

Where there are simultaneous changes in more than one constraint RHS values, the 100 per cent Rule must be applied.

Reduced Cost

$$\text{Reduced cost / unit of activity} = \left(\begin{array}{l} \text{Cost of consumed} \\ \text{resources per unit} \\ \text{of activity} \end{array} \right) - \left(\begin{array}{l} \text{Profit per unit} \\ \text{of activity} \end{array} \right)$$

If the activity's reduced cost per unit is positive, then its unit cost of consumed resources is higher than its unit profit, and the activity should be discarded. This means that the value of its associated variable in the optimum solution should be zero.

Alternatively, an activity that is economically attractive will have a zero reduced cost in the optimum solution signifying equilibrium between the output (unit profit) and the input (unit cost of consumed resources).

In the problem, both x_1 and x_2 assume positive values in the optimum solution and hence have zero reduced cost.

Considering one more variable x_3 with profit Rs. 50

$$Z_{\max} = 40x_1 + 30x_2 + 50x_3$$

Subject to constraints,

$$2x_2 + 1x_3 \leq 50 \quad \dots \dots \dots \text{(ii)}$$

$$6x_1 + 3x_2 + 3x_3 \leq 150 \quad \dots \dots \dots \text{(iii)}$$

where $x_1, x_2, x_3 \geq 0$

The sensitivity analysis of the problem is shown in the computer output below in Table 5.12.

Table 5.12: Sensitivity Analysis

The screenshot shows the TOPS software interface with the following details:

TOPS Output File

LINEAR PROGRAMMING

LINEAR PROGRAMMING OUTPUT SUMMARY

Title: test
Final Iteration No.: 3
Objective Value (Max) = 150.56

Variable	Value	Obj. Cost	Obj. Min. Cost
x1	15.63	49.00	625.00
x2	8.00	30.00	8.00
x3	18.75	50.00	827.50
Constraint	Value	Obj. Min. Cost	
S1	175.00	0.00	
S2	58.00	31.75	
S3	118.00	0.00	

Sensitivity Analysis

Variable	Current Obj. Cost	Obj. Obj. Cost	Obj. Min. Cost	Reduced Cost
x1	49.00	33.33	198.00	9.00
x2	30.00	infinity	625.00	12.50
x3	50.00	33.33	50.00	0.00
Constraint	Obj. Min. Cost	Obj. Min. Cost	Reduced Cost	
S1	175.00	160.00	300.00	7.50
S2	58.00	38.75	infinity	0.00
S3	118.00	87.50	362.50	5.00

Buttons at the bottom:

- ViewModif Input Data
- Addl. Mems
- Edit TOPS

The reduced cost indicates how much the objective function co-efficient for a particular variable would have to improve before that decision function assumes a positive value in the optimal solution.

The reduced cost of Rs.12.50 for decision variable x_2 tells us that the profit contribution would have to increase to at least $30 + 12.50 = 42.50$ before x_3 could assume a positive value in the optimal solution.

Check Your Progress 5.2

1. What is Duality concept?
2. What is meant by degeneracy in Linear Programming?

Notes: (a) Write your answer in the space given below.
(b) Please go through the lesson sub-head thoroughly you will get your answers in it.
(c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

5.12 LET US SUM UP

Thus LP is a planning technique of selecting the best possible (optimal) strategy among number of alternatives. The chosen strategy is said to be the best because it involves minimization/maximization of some desired action e.g. maximization of profits, minimization of costs, smoothening running of the business.

5.13 LESSON-END ACTIVITIES

1. Linear Programming is a general method usable for a wide range of problems.
Go to any nutrition center which sells health-food. Bring into play the applications of LP in formation and building.
2. LP is no doubt an vital problem. Not in this counters of petty problems with only a couple of variables, but is much bigger problem.
Exaggerate this logic with the help of illustrations which can be matched and linked with you real-life-situations.

5.14 KEYWORDS

Slack

Simplex method

Surplus

5.15 QUESTIONS FOR DISCUSSION

1. **Write True or False against each statement:**
 - (a) Artificial variable are imaginary and do not have any physical meaning.
 - (b) Simplex method solve the LPP in iteration to enhance the value of the objective function.
 - (c) Sensitivity analysis can not be used to provide information and to determine solution with these changes.

2. **Briefly comment on the following statement:**
 - (a) Two or more entries in the ratio column.
 - (b) LP is a planning techniques.
 - (c) LP techniques are used to optimise the resources for best result.
 - (d) LP is a part of management science.
 - (e) Algebraic techniques is used to solve large problems using simplex method.

5.16 TERMINAL QUESTIONS

1. Explain the procedure involved in the simplex method to determine the optimum solution.
2. What are slack, surplus and artificial variables ?
3. What is degeneracy in LP problems ? When does it occur ? How can degeneracy problem be resolved ?
4. What is a basic variable and a non-basic variable ?
5. Explain what is an unbounded solution in LPP.
6. Differentiate between primal and dual problems.
7. Why is the simplex method more advantageous than the graphical method?
8. What are the rules in selecting key column, key row and pivotal element?
9. Discuss the role of sensitivity analysis in linear programming.
10. In sensitivity analysis, explain
 - i. The effect of change of objective function coefficients
 - ii. The effect of change in the right hand side of constraints

Exercise Problems

1. A company manufactures three products A, B and C, which require three raw materials I, II and III. The table given below shows the amount of raw materials required per kg of each product. The resource availability per day and the profit contribution for each product is also given.

Product	A	B	C	Availability (kg)
Raw Material				
I	4	1	6	800
II	5	6	8	1500
III	2	4	1	1200
Profit per unit (Rs)	9	10	6	

- i. Formulate the problem as a linear programming problem.
 - ii. Solve the problem and determine the optimal product mix.

2. A metal fabricator manufactures three types of windows. Each of the windows needs four processes. The time taken on various machines differ due to the size of windows. The time taken and available hours are given in the table below:

Window Type	Cutting	Heat Treating	Forging	Grinding
A	5	7	1	4
B	7	4	4	8
C	4	8	6	2
Available time (Hrs)	20	24	28	22

The profit contribution for windows A, B and C are Rs. 3.00, Rs. 4.00 and Rs. 5.00 respectively.

- a. Formulate the problem.
 - b. Solve the problem using simplex method to maximize the profit.
 - c. Determine the excess time available in each processes and by how much.

3. Solve the following LPP using simplex method.

$$\text{Maximize } , Z = 2x_1 + x_2$$

Subject to constraints,

$$4x_1 - x_2 \leq 8 \quad \dots \dots \dots \text{(iii)}$$

where

$$x_1, x_2 \geq 0$$

4. Solve the following LPP:

$$Z_{\max} = 20x_1 + 28x_2 + 23x_3$$

Subject to constraints,

$$2x_1 + x_2 + 2x_3 \leq 100 \quad \dots \dots \dots \text{(ii)}$$

$$3x_1 + 2x_2 + x_3 \leq 50 \quad \dots \dots \dots \text{(iii)}$$

where

$$x_1, x_2, x_3 \geq 0$$

5. Three high precision products are manufactured by a Hi-Tech Machine Tools Company. All the products must undergo process through three machining centers A, B and C. The machine hours required per unit are,

Machining Center	Product		
	I	II	III
A	2	4	6
B	3	6	2
C	3	2	1

The available time in machine hours per week is

Machining Center	Machine Hours Per Week
A	150
B	100
C	120

It is estimated that the unit profits of the product are

Product	Unit Profits (Rs)
I	3
II	4
III	6

- a. Formulate the problem as a LPP.
- b. Solve the problem to determine the optimal solution. What is the number of units to be made on each product.
- c. Does machining center C has any extra time to spare? If so, how much spare time is available ?
- d. If additional 10 machine hours are available with machining center A, then what is the optimal product mix ? What is the change in the value of profit ?
6. Raghu Constructions is considering four projects over the next 3 years. The expected returns of each project and cash outlays for these projects are listed in the tables given. All values are in Lacs of Rupees.

Project	Cash outlay (lakh Rs.)			Return
	Year 1	Year 2	Year 3	
1	12.32	11.10	9.50	42.25
2	11.15	9.75	8.11	31.20
3	7.65	5.50	4.75	15.10
4	10.71	10.31	7.77	12.05
Available funds (lakh Rs.)	110.00	40.00	35.00	

Raghu has to decide to undertake construction projects. Ignore the time value of money. As a consultant, what suggestion you would like to give Raghu in deciding about the projects to select. Determine the solution using TORA.

7. Solve the following LP Problem using Big M Method.

$$\text{Minimize, } Z = 2x_1 + 9x_2 + x_3$$

Subject to constraints,

$$x_1 + 4x_2 + 2x_3 \geq 5 \quad \dots \dots \dots \text{(i)}$$

$$3x_1 + x_2 + 2x_3 \geq 4 \quad \dots \dots \dots \text{(ii)}$$

$$\text{where } x_1, x_2, x_3 \geq 0$$

8. Solve the following LPP

$$Z_{\min} = 4x_1 + x_2$$

Unit-II

LESSON

6

TRANSPORTATION MODEL

CONTENTS

- 6.0 Aims and Objectives
- 6.1 Introduction
- 6.2 Mathematical Formulation
- 6.3 Network Representation of Transportation Model
- 6.4 General Representation of Transportation Model
- 6.5 Use of Linear Programming to Solve Transportation Problem
- 6.6 Formulation of LP model
- 6.7 Solving Transportation Problem Using Computer
- 6.8 Balanced Transportation Problem
- 6.9 Unbalanced Transportation Problem
- 6.10 Procedure to Solve Transportation Problem
- 6.11 Degeneracy in Transportation Problems
- 6.12 Maximization Transportation Problem
- 6.13 Prohibited Routes Problem
- 6.14 Transhipment Problem
- 6.15 Let us Sum Up
- 6.16 Lesson-end Activity
- 6.17 Keywords
- 6.18 Questions for Discussion
- 6.19 Terminal Questions
- 6.20 Model Answers to Questions for Discussion
- 6.21 Suggested Readings

6.0 AIMS AND OBJECTIVES

In this unit we would be able to learn the Time Management Models. i.e. Transportation and Assignment Models, thus would be able to learn transportation models in this lesson and also we will talk about transhipment problems.

6.1 INTRODUCTION

Transportation problem is a particular class of linear programming, which is associated with day-to-day activities in our real life and mainly deals with logistics. It helps in solving problems on distribution and transportation of resources from one place to another. The goods are transported from a set of sources (e.g., factory) to a set of destinations (e.g., warehouse) to meet the specific requirements. In other words, transportation problems deal with the transportation of a product manufactured at different plants (*supply origins*) to a number of different warehouses (*demand destinations*). The objective is to satisfy the demand at destinations from the supply constraints at the minimum transportation cost possible. To achieve this objective, we must know the quantity of available supplies and the quantities demanded. In addition, we must also know the location, to find the cost of transporting one unit of commodity from the place of origin to the destination. The model is useful for making strategic decisions involved in selecting optimum transportation routes so as to allocate the production of various plants to several warehouses or distribution centers.

The transportation model can also be used in making location decisions. The model helps in locating a new facility, a manufacturing plant or an office when two or more number of locations is under consideration. The total transportation cost, distribution cost or shipping cost and production costs are to be minimized by applying the model.

6.2 MATHEMATICAL FORMULATION

The transportation problem applies to situations where a single commodity is to be transported from various sources of supply (**origins**) to various demands (**destinations**).

Let there be m sources of supply S_1, S_2, \dots, S_m having a_i ($i = 1, 2, \dots, m$) units of supplies respectively to be transported among n destinations D_1, D_2, \dots, D_n with b_j ($j = 1, 2, \dots, n$) units of requirements respectively. Let C_{ij} be the cost for shipping one unit of the commodity from source i , to destination j for each route. If x_{ij} represents the units shipped per route from source i , to destination j , then the problem is to determine the transportation schedule which minimizes the total transportation cost of satisfying supply and demand conditions.

The transportation problem can be stated mathematically as a linear programming problem as below:

$$\text{Minimize } Z = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

Subject to constraints,

$$\sum_{j=1}^n x_{ij} = a_i, \quad i = 1, 2, \dots, m \text{ (supply constraints)}$$

$$\sum_{i=1}^m x_{ij} = b_j, \quad j = 1, 2, \dots, n \text{ (demand constraints)}$$

and $x_{ij} \geq 0$ for all $i = 1, 2, \dots, m$ and,

$$j = 1, 2, \dots, n$$

6.3 NETWORK REPRESENTATION OF TRANSPORTATION MODEL

The transportation model is represented by a network diagram in Figure 6.1.

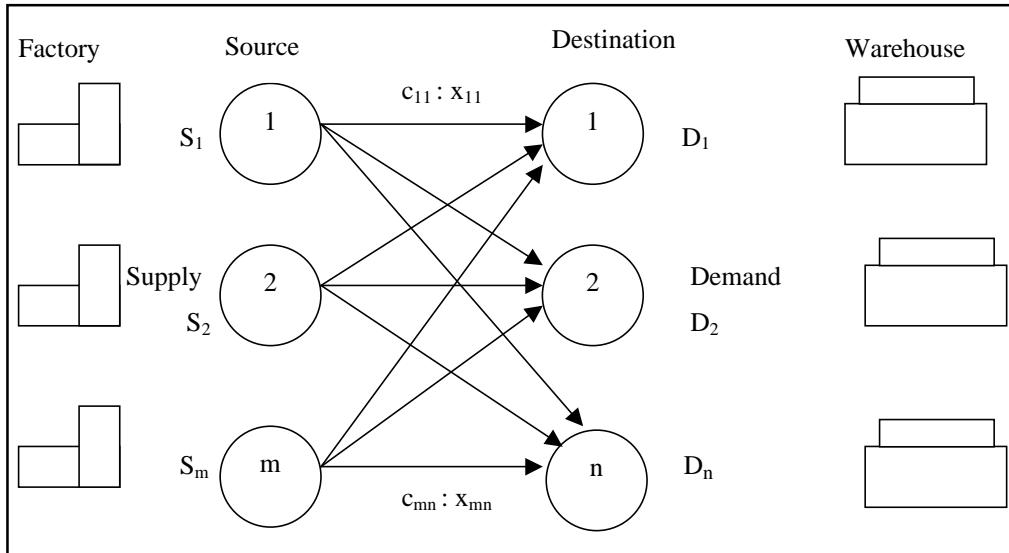


Figure 6.1: Network Transportation Model

where,

m be the number of sources,

n be the number of destinations,

S_m be the supply at source m ,

D_n be the demand at destination n ,

c_{ij} be the cost of transportation from source i to destination j , and

x_{ij} be the number of units to be shipped from source i to destination j .

The objective is to minimize the total transportation cost by determining the unknowns x_{ij} , i.e., the number of units to be shipped from the sources and the destinations while satisfying all the supply and demand requirements.

6.4 GENERAL REPRESENTATION OF TRANSPORTATION MODEL

The Transportation problem can also be represented in a tabular form as shown in Table 6.1

Let C_{ij} be the cost of transporting a unit of the product from i^{th} origin to j^{th} destination.

a_i be the quantity of the commodity available at source i ,

b_j be the quantity of the commodity needed at destination j , and

x_{ij} be the quantity transported from i^{th} source to j^{th} destination

Table 6.1: Tabular Representation of Transportation Model

<i>To</i> <i>From</i>	D_1	D_2	...	D_n	<i>Supply</i> A_i
S_1	C_{11} x_{11}	C_{12} x_{12}	...	C_{1n}	A_1
S_2	C_{21} x_{21}	C_{22} x_{22}	...	C_{2n}	A_2
.
S_m	C_{m1} x_{m1}	C_{m2} x_{m2}	...	C_{mn}	A_m
B_j	B_1	B_2	...	B_n	$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$$

If the total supply is equal to total demand, then the given transportation problem is a balanced one.

6.5 USE OF LINEAR PROGRAMMING TO SOLVE TRANSPORTATION PROBLEM

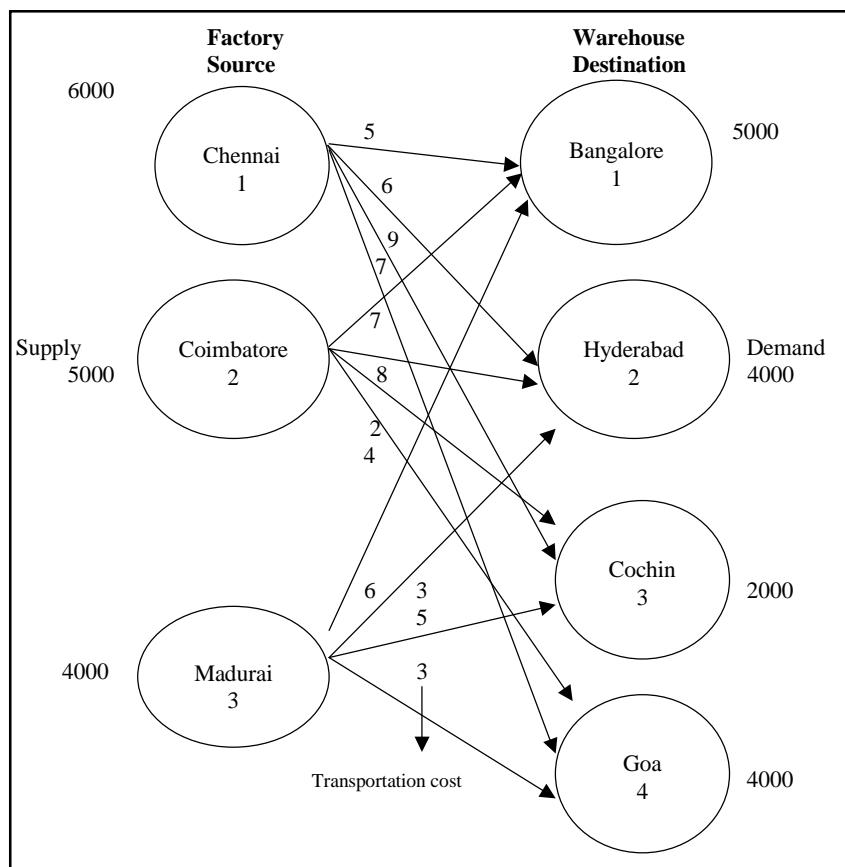


Figure 6.2: Linear Programming Solution

The network diagram shown in Figure 6.2 represents the transportation model of M/s GM Textiles units located at Chennai, Coimbatore and Madurai. GM Textiles produces ready-made garments at these locations with capacities 6000, 5000 and 4000 units per week at Chennai, Coimbatore and Madurai respectively. The textile unit distributes its ready-made garments through four of its wholesale distributors situated at four locations Bangalore, Hyderabad, Cochin and Goa. The weekly demand of the distributors are 5000, 4000, 2000 and 4000 units for Bangalore, Hyderabad, Cochin and Goa respectively.

The cost of transportation per unit varies between different supply points and destination points. The transportation costs are given in the network diagram.

The management of GM Textiles would like to determine the number of units to be shipped from each textile unit to satisfy the demand of each wholesale distributor. The supply, demand and transportation cost are as follows:

Table 6.2: Production Capacities

Supply	Textile Unit	Weekly Production (Units)
1	Chennai	6000
2	Coimbatore	5000
3	Madurai	4000

Table 6.3: Demand Requirements

Destination	Wholesale Distributor	Weekly Demand (Units)
1	Bangalore	5000
2	Hyderabad	4000
3	Cochin	2000
4	Goa	4000

Table 6.4: Transportation cost per unit

Supply	Destination			
	B'lore	Hyderabad	Cochin	Goa
Chennai	5	6	9	7
Coimbatore	7	8	2	4
Madurai	6	3	5	3

A linear programming model can be used to solve the transportation problem.

Let,

X_{11} be number of units shipped from source1 (Chennai) to destination 1 (B'lore).

X_{12} be number of units shipped from source1 (Chennai) to destination 2 (Hyderabad).

X_{13} number of units shipped from source 1 (Chennai) to destination 3 (Cochin).

X_{14} number of units shipped from source 1 (Chennai) to destination 4 (Goa)

and so on.

X_{ij} = number of units shipped from source i to destination j, where $i = 1, 2, \dots, m$

and,

$j = 1, 2, \dots, n$.

Check Your Progress 6.1

1. What is the transportation problem?
2. Give a tabular representation of transportation model.

Contd....

- Notes:**

 - (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Chek Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

6.6 FORMULATION OF LP MODEL

Objective function: The objective is to minimize the total transportation cost. Using the cost data table, the following equation can be arrived at:

Transportation cost for units

$$\text{shipped from Chennai} = 5x_{11} + 6x_{12} + 9x_{13} + 7x_{14}$$

Transportation cost for units

shipped from Coimbatore = $7x_{21} + 8x_{22} + 2x_{23} + 4x_{24}$

Transportation cost for units

$$\text{shipped from Madurai} = 6x_{31} + 3x_{32} + 5x_{33} + 3x_{34}$$

Combining the transportation cost for all the units shipped from each supply point with the objective to minimize the transportation cost, the objective function will be,

$$\text{Minimize } Z = 5x_{11} + 6x_{12} + 9x_{13} + 7x_{14} + 7x_{21} + 8x_{22} + 2x_{23} + 4x_{24} + 6x_{31} + 3x_{32} + 5x_{33} + 3x_{34}$$

Constraints:

In transportation problems, there are supply constraints for each source, and demand constraints for each destination.

Supply constraints:

For Chennai, $x_{11} + x_{12} + x_{13} + x_{14} \leq 6000$

For Coimbatore, $x_{21} + x_{22} + x_{23} + x_{24} \leq 5000$

For Madurai, $x_{21} + x_{22} + x_{23} + x_{24} \leq 4000$

Demand constraints:

For B'lore, $x_{11} + x_{21} + x_{31} = 5000$

For Hyderabad, $x_{12} + x_{23} + x_{31} = 4000$

For Cochin, $x_{12} + x_{22} + x_{32} = 2000$

For Goa, $x_1 + x_2 + x_3 = 4000$

The linear programming model for GM Textiles will be write in the next line. Minimize $Z = 5x_{11} + 6x_{12} + 9x_{21} + 7x_{22} + 7x_{31} + 8x_{32} + 2x_{41} + 4x_{42} + 6x_{51} + 3x_{52} + 5x_{61} + 3x_{62}$

Subject to constraints

$$X_1 + X_2 + X_3 + X_4 \leq 6000 \quad (i)$$

- $X_{21} + X_{22} + X_{23} + X_{24} \leq 5000$ (ii) Transportation Model
 $X_{31} + X_{32} + X_{33} + X_{34} \leq 4000$ (iii)
 $X_{11} + X_{21} + X_{31} = 5000$ (iv)
 $X_{12} + X_{22} + X_{32} = 4000$ (v)
 $X_{13} + X_{23} + X_{33} = 2000$ (vi)
 $X_{14} + X_{24} + X_{34} = 4000$ (vii)

Where, $x_{ij} \geq 0$ for $i = 1, 2, 3$ and $j = 1, 2, 3, 4$.

6.7 SOLVING TRANSPORTATION PROBLEM USING COMPUTER

Input screen for solving TP & LP models using TORA

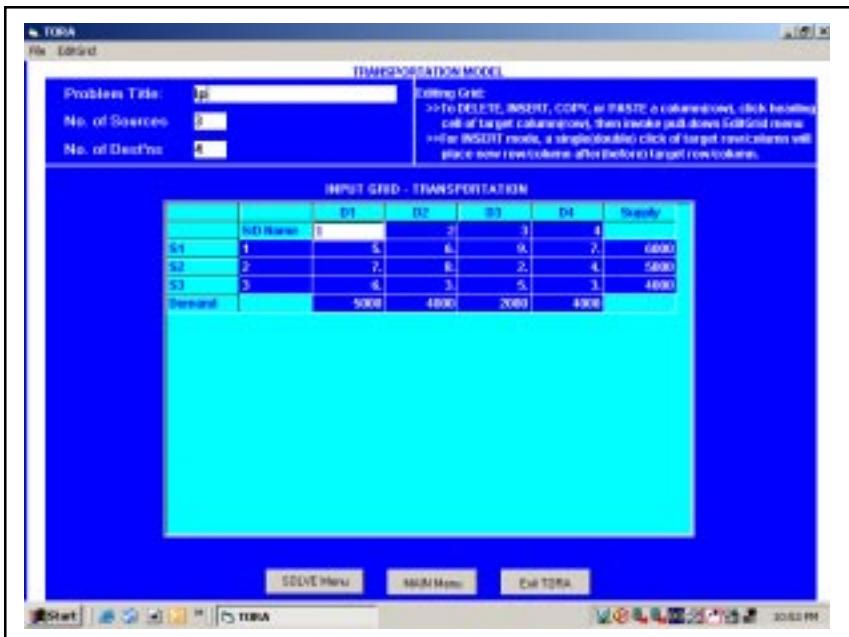
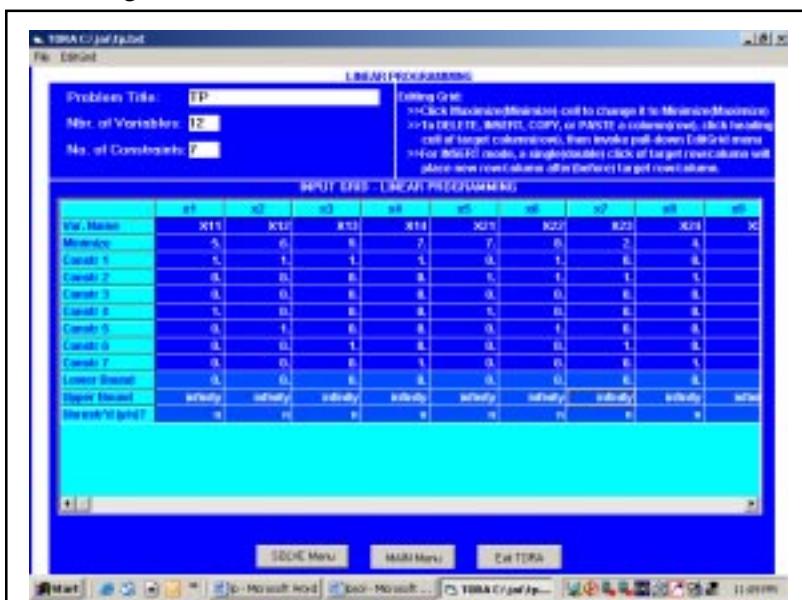


Figure 6.3: TORA Screen for TP Model

Output screen using TP & LP models



Example 1: Consider the following transportation problem (Table 6.5) and develop a linear programming (LP) model.

Table 6.5: Transportation Problem

Source	Destination			
	1	2	3	Supply
1	15	20	30	350
2	10	9	15	200
3	14	12	18	400
Demand	250	400	300	

Solution: Let x_{ij} be the number of units to be transported from the source i to the destination j , where $i = 1, 2, 3, \dots, m$ and $j = 1, 2, 3, \dots, n$.

The linear programming model is

$$\text{Minimize } Z = 15x_{11} + 20x_{12} + 30x_{13} + 10x_{21} + 9x_{22} + 15x_{23} + 14x_{31} + 12x_{32} + 18x_{33}$$

Subject to constraints,

$$x_{11} + x_{12} + x_{13} \leq 350 \quad \dots \dots \dots \text{(i)}$$

$$x_{31} + x_{32} + x_{33} \leq 400 \quad \dots \dots \dots \text{(iii)}$$

$$x_{11} + x_{12} + x_{21} = 250 \quad \dots \dots \dots \text{(iv)}$$

$x_{ij} \geq 0$ for all i and j.

In the above LP problem, there are $m \times n = 3 \times 3 = 9$ decision variables and $m + n = 3 + 3 = 6$ constraints.

6.8 BALANCED TRANSPORTATION PROBLEM

When the total supplies of all the sources are equal to the total demand of all destinations, the problem is a ***balanced transportation problem***.

Total supply = Total demand

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$$

The problem given in Example 3.1 represents a balanced transportation problem.

6.9 UNBALANCED TRANSPORTATION PROBLEM

When the total supply of all the sources is not equal to the total demand of all destinations, the problem is an ***unbalanced transportation problem***.

Total supply¹ Total demand

$$\sum_{i=1}^m a_i \leq \sum_{j=1}^n b_j$$

Demand Less than Supply

In real-life, supply and demand requirements will rarely be equal. This is because of variation in production from the supplier end, and variations in forecast from the customer

end. Supply variations may be because of shortage of raw materials, labour problems, improper planning and scheduling. Demand variations may be because of change in customer preference, change in prices and introduction of new products by competitors. These unbalanced problems can be easily solved by introducing **dummy sources** and **dummy destinations**. If the total supply is greater than the total demand, a dummy destination (**dummy column**) with demand equal to the supply surplus is added. If the total demand is greater than the total supply, a dummy source (dummy row) with supply equal to the demand surplus is added. The unit transportation cost for the dummy column and dummy row are assigned zero values, because no shipment is actually made in case of a dummy source and dummy destination.

Example 2: Check whether the given transportation problem shown in Table 6.6 is a balanced one. If not, convert the unbalanced problem into a balanced transportation problem.

Table 6.6: Transportation Model with Supply Exceeding Demand

Source	Destination			Supply
	1	2	3	
1	25	45	10	200
2	30	65	15	100
3	15	40	55	400
Demand	200	100	300	

Solution: For the given problem, the total supply is not equal to the total demand.

$$\sum_{i=1}^3 a_i = \sum_{j=1}^3 b_j$$

since,

$$\sum_{i=1}^3 a_i = 700 \text{ and } \sum_{j=1}^3 b_j = 600$$

The given problem is an unbalanced transportation problem. To convert the unbalanced transportation problem into a balanced problem, add a dummy destination (dummy column).

i.e., the demand of the dummy destination is equal to,

$$\sum_{i=1}^3 a_i - \sum_{j=1}^3 b_j$$

Thus, a dummy destination is added to the table, with a demand of 100 units. The modified table is shown in Table 6.7 which has been converted into a balanced transportation table. The unit costs of transportation of dummy destinations are assigned as zero.

Table 6.7: Dummy Destination Added

Source	Destination				Supply
	1	2	3	4	
1	25	45	10	0	200
2	30	65	15	0	100
3	15	40	55	0	400
Demand	200	100	300	100	700/700

Similarly,

If $\sum_{j=1}^n b_j > \sum_{i=1}^m a_i$ then include a dummy source to supply the excess demand.

Demand Greater than Supply

Example 3: Convert the transportation problem shown in Table 6.8 into a balanced problem.

Table 6.8: Demand Exceeding Supply

Source	Destination				Supply
	1	2	3	4	
1	10	16	9	12	200
2	12	12	13	5	300
3	14	8	13	4	300
Demand	100	200	450	250	1000/800

Solution: The given problem is,

$$\sum_{j=1}^4 b_j > \sum_{i=1}^3 a_i$$

$$\sum_{i=1}^3 a_i = 800 \text{ and } \sum_{j=1}^4 b_j = 1000$$

The given problem is an unbalanced one. To convert it into a balanced transportation problem, include a dummy source (dummy row) as shown in Table 6.9

Table 6.9: Balanced TP Model

Source	Destination				Supply
	1	2	3	4	
1	10	16	9	12	200
2	12	12	13	5	300
3	14	8	13	4	300
4	0	0	0	0	200
Demand	100	200	450	250	1000/1000

6.10 PROCEDURE TO SOLVE TRANSPORTATION PROBLEM

Step 1: Formulate the problem.

Formulate the given problem and set up in a matrix form. Check whether the problem is a balanced or unbalanced transportation problem. If unbalanced, add dummy source (row) or dummy destination (column) as required.

Step 2: Obtain the initial feasible solution.

The initial feasible solution can be obtained by any of the following three methods:

- i. Northwest Corner Method (NWC)
- ii. Least Cost Method (LCM)
- iii. Vogel's Approximation Method (VAM)

The transportation cost of the initial basic feasible solution through Vogel's approximation method, VAM will be the least when compared to the other two methods which gives the value nearer to the optimal solution or optimal solution itself. Algorithms for all the three methods to find the initial basic feasible solution are given.

Algorithm for North-West Corner Method (NWC)

Transportation Model

- (i) Select the North-west (i.e., upper left) corner cell of the table and allocate the maximum possible units between the supply and demand requirements. During allocation, the transportation cost is completely discarded (not taken into consideration).
- (ii) Delete that row or column which has no values (fully exhausted) for supply or demand.
- (iii) Now, with the new reduced table, again select the North-west corner cell and allocate the available values.
- (iv) Repeat steps (ii) and (iii) until all the supply and demand values are zero.
- (v) Obtain the initial basic feasible solution.

Algorithm for Least Cost Method (LCM)

- (i) Select the smallest transportation cost cell available in the entire table and allocate the supply and demand.
- (ii) Delete that row/column which has exhausted. The deleted row/column must not be considered for further allocation.
- (iii) Again select the smallest cost cell in the existing table and allocate. (Note: In case, if there are more than one smallest costs, select the cells where maximum allocation can be made)
- (iv) Obtain the initial basic feasible solution.

Algorithm for Vogel's Approximation Method (VAM)

- (i) Calculate penalties for each row and column by taking the difference between the smallest cost and next highest cost available in that row/column. If there are two smallest costs, then the penalty is zero.
- (ii) Select the row/column, which has the largest penalty and make allocation in the cell having the least cost in the selected row/column. If two or more equal penalties exist, select one where a row/column contains minimum unit cost. If there is again a tie, select one where maximum allocation can be made.
- (iii) Delete the row/column, which has satisfied the supply and demand.
- (iv) Repeat steps (i) and (ii) until the entire supply and demands are satisfied.
- (v) Obtain the initial basic feasible solution.

Remarks: The initial solution obtained by any of the three methods must satisfy the following conditions:

- (a) The solution must be feasible, i.e., the supply and demand constraints must be satisfied (also known as rim conditions).
- (b) The number of positive allocations, N must be equal to $m+n-1$, where m is the number of rows and n is the number of columns.

6.11 DEGENERACY IN TRANSPORTATION PROBLEMS

Step 3: Check for degeneracy

The solution that satisfies the above said conditions $N = m + n - 1$ is a non-degenerate basic feasible solution otherwise, it is a degenerate solution. Degeneracy may occur either at the initial stage or at subsequent iterations.

If number of allocations, $N = m + n - 1$, then degeneracy does not exist. Go to Step 5.

If number of allocations, $N \neq m + n - 1$, then degeneracy does exist. Go to Step 4.

Step 4: Resolving degeneracy

To resolve degeneracy at the initial solution, allocate a small positive quantity e to one or more unoccupied cell that have lowest transportation costs, so as to make $m + n - 1$ allocations (i.e., to satisfy the condition $N = m + n - 1$). The cell chosen for allocating e must be of an independent position. In other words, the allocation of e should avoid a closed loop and should not have a path.

The following Table 6.10 shows independent allocations.

Table 6.10: Independent Allocations

*	*	*	
	*		

*	*		
		*	
	*		*

The following Tables 6.10 (a), (b) and (c) show non-independent allocations.

Table 6.10 (a): Non-Independent Allocations

		*	*
	*	*	

Table 6.10 (b)

	*		*
	*		*

Table 6.10 (c)

*	*		
	*		
*	*		

Optimal Solution

Step 5: Test for optimality

The solution is tested for optimality using the Modified Distribution (MODI) method (also known as U-V method).

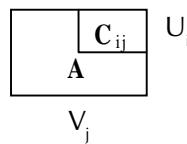
Once an initial solution is obtained, the next step is to test its optimality. An optimal solution is one in which there are no other transportation routes that would reduce the total transportation cost, for which we have to evaluate each unoccupied cell in the table in terms of opportunity cost. In this process, if there is no negative opportunity cost, and the solution is an optimal solution.

- (i) Row 1, row 2,..., row i of the cost matrix are assigned with variables U_1, U_2, \dots, U_i and the column 1, column 2,..., column j are assigned with variables V_1, V_2, \dots, V_j respectively.

- (ii) Initially, assume any one of U_i values as zero and compute the values for U_1, U_2, \dots, U_i and V_1, V_2, \dots, V_j by applying the formula for occupied cell.

For occupied cells,

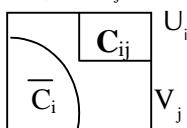
$$C_{ij} + U_i + V_j = 0$$



- (iii) Obtain all the values of C_{ij} for unoccupied cells by applying the formula for unoccupied cell.

For unoccupied cells,

$$\text{Opportunity Cost, } \bar{C}_{ij} = C_{ij} + U_i + V_j$$



If \bar{C}_{ij} values are > 0 then, the basic initial feasible solution is optimal.

Go to step 7.

If \bar{C}_{ij} values are $= 0$ then, the multiple basic initial feasible solution exists. Go to step 7.

If \bar{C}_{ij} values are < 0 then, the basic initial feasible solution is not optimal. Go to step 6.

Step 6: Procedure for shifting of allocations

Select the cell which has the most negative \bar{C}_{ij} value and introduce a positive quantity called ‘q’ in that cell. To balance that row, allocate a ‘-q’ to that row in occupied cell. Again, to balance that column put a positive ‘q’ in an occupied cell and similarly a ‘-q’ to that row. Connecting all the ‘q’s and ‘-q’s, a closed loop is formed.

Two cases are represented in Table 6.11(a) and 6.11(b). In Table 6.11(a) if all the q allocations are joined by horizontal and vertical lines, a closed loop is obtained.

The set of cells forming a closed loop is

$$CL = \{(A, 1), (A, 3), (C, 3), (C, 4), (E, 4), (E, 1), (A, 1)\}$$

The loop in Table 6.11(b) is not allowed because the cell (D3) appears twice.

Table 6.11(a): Showing Closed Loop

	1	2	3	4
A	*		*	
B				
C			*	*
D				
E	*			*

Table 6.11(b)

	1	2	3	4
A	*			
B	*		*	
C				
D	*		*	*
E			*	*

Conditions for forming a loop

- (i) The start and end points of a loop must be the same.
- (ii) The lines connecting the cells must be horizontal and vertical.
- (iii) The turns must be taken at occupied cells only.
- (iv) Take a shortest path possible (for easy calculations).

Remarks on forming a loop

- (i) Every loop has an even number of cells and at least four cells
- (ii) Each row or column should have only one '+' and '-' sign.
- (iii) Closed loop may or may not be square in shape. It can also be a rectangle or a stepped shape.
- (iv) It doesn't matter whether the loop is traced in a clockwise or anti-clockwise direction.

Take the most negative ' $-q$ ' value, and shift the allocated cells accordingly by adding the value in positive cells and subtracting it in the negative cells. This gives a new improved table. Then go to step 5 to test for optimality.

Step 7: Calculate the Total Transportation Cost.

Since all the \bar{C}_{ij} values are positive, optimality is reached and hence the present allocations are the optimum allocations. Calculate the total transportation cost by summing the product of allocated units and unit costs.

Example 4: The cost of transportation per unit from three sources and four destinations are given in Table 6.12. Obtain the initial basic feasible solutions using the following methods.

- (i) North-west corner method
- (ii) Least cost method
- (iii) Vogel's approximation method

Table 6.12: Transportation Model

Source	Destination				Supply
	1	2	3	4	
1	4	2	7	3	250
2	3	7	5	8	450
3	9	4	3	1	500
Demand	200	400	300	300	1200

Solution: The problem given in Table 6.13 is a balanced one as the total sum of supply is equal to the total sum of demand. The problem can be solved by all the three methods.

North-West Corner Method: In the given matrix, select the North-West corner cell. The North-West corner cell is (1,1) and the supply and demand values corresponding to cell (1,1) are 250 and 200 respectively. Allocate the maximum possible value to satisfy the demand from the supply. Here the demand and supply are 200 and 250 respectively. Hence allocate 200 to the cell (1,1) as shown in Table 6.13.

Table 6.13: Allocated 200 to the Cell (1, 1)

		Destination				Supply	
Source		1	2	3	4		
		200	4	2	7	3	
		3	7	5	8	450	
		9	4	3	1	500	
Demand		200	400	300	300		
		0					

Now, delete the exhausted column 1 which gives a new reduced table as shown in Tables 6.14 (a, b, c, d). Again repeat the steps.

Table 6.14 (a): Exhausted Column 1 Deleted

		Destination			Supply	
Source		2	3	4		
		50	7	3	3	
		7	5	8	450	
		400	300	350	500	
Demand		350				

Table after deleting Row 1

Table 6.14 (b): Exhausted Row 1 Deleted

		Destination			Supply	
Source		2	3	4		
		350	5	8	100	
		4	3	1	500	
		350	300	300		
Demand		0				

Table after deleting column 2

Table 6.14 (c): Exhausted Column 2 Deleted

		Destination		Supply	
Source		3	4		
		100	8	100	
		3	1	500	
		300	300		
Demand		200			

Finally, after deleting Row 2, we have

Table 6.14 (d): Exhausted Row 2 Deleted

		Destination		Supply
Source	3	3	4	
		300	200	
Demand		300	200	
		0	0	

Now only source 3 is left. Allocating to destinations 3 and 4 satisfies the supply of 500. The initial basic feasible solution using North-west corner method is shown in Table 6.15

Table 6.15: Initial Basic Feasible Solution Using NWC Method

		Plant					Supply
Warehouse	1	1	2	3	4	5	
1	10	2	3	25	15	9	25
2	20				10		30
3	5	10	15	2	4		20
4		20					30
	15	5	14	7	15		
	20	15	13	5	8	25	105
Demand	20	20	30	10	25	105	

$$\begin{aligned}
 \text{Transportation cost} &= (4 \times 200) + (2 \times 50) + (7 \times 350) + (5 \times 100) + \\
 &\quad (2 \times 300) + (1 \times 300) \\
 &= 800 + 100 + 2450 + 500 + 600 + 300 \\
 &= \text{Rs. 4,750.00}
 \end{aligned}$$

Least Cost Method

Select the minimum cost cell from the entire Table 6.16, the least cell is (3,4). The corresponding supply and demand values are 500 and 300 respectively. Allocate the maximum possible units. The allocation is shown in Table 6.16.

Table 6.16: Allocation of Maximum Possible Units

		Destination				Supply
Source	1	1	2	3	4	
1	4	2	7	3		250
2	3	7	5	8		450
3	9	4	3	1		500
				300		200
Demand	200	400	300	300	0	

From the supply value of 500, the demand value of 300 is satisfied. Subtract 300 from the supply value of 500 and subtract 300 from the demand value of 300. The demand of

destination 4 is fully satisfied. Hence, delete the column 4; as a result we get, the table as shown in Table 6.17.

Table 6.17: Exhausted Column 4 Deleted

		Destination			Supply
		1	2	3	
Source	1	4	2	7	250
	2	3	7	5	450
	3	9	4	3	200
Demand		200	400	300	
			150		

Now, again take the minimum cost value available in the existing table and allocate it with a value of 250 in the cell (1,2).

The reduced matrix is shown in Table 6.18

Table 6.18: Exhausted Row 1 Deleted

		Desitnation			Supply
		1	2	3	
Source	2	3	7	5	250
	3	9	4	3	450
Demand		200	150	300	

In the reduced Table 6.18, the minimum value 3 exists in cell (2,1) and (3,3), which is a tie. If there is a tie, it is preferable to select a cell where maximum allocation can be made. In this case, the maximum allocation is 200 in both the cells. Choose a cell arbitrarily and allocate. The cell allocated in (2,1) is shown in Table 6.18. The reduced matrix is shown in Table 6.19.

Table 6.19: Reduced Matrix

		Destination			Supply
		2	3		
Source	2	7	5	250	
	3	4	3	200	0
Demand		150	300	100	

Now, deleting the exhausted demand row 3, we get the matrix as shown in Table 6.20

Table 6.20: Exhausted Row 3 Deleted

		Destination			Supply
		2	3		
Source	2	7	5	250	0
	3	4	3	200	0
Demand		150	100	0	

The initial basic feasible solution using least cost method is shown in a single Table 6.21

Table 6.21: Initial Basic Feasible Solution Using LCM Method

		Destination				Supply
		1	2	3	4	
Source	1	4 250	2 250	7	3	250
	2	3 200	7 150	5 100	8	450
	3	9	4	3 200	1 300	500
Demand		200	400	300	300	0

$$\begin{aligned}
 \text{Transportation Cost} &= (2 \times 250) + (3 \times 200) + (7 \times 150) + (5 \times 100) + (3 \times 200) + \\
 &\quad (1 \times 300) \\
 &= 500 + 600 + 1050 + 500 + 600 + 300 = \text{Rs. 3550}
 \end{aligned}$$

Vogel's Approximation Method (VAM): The penalties for each row and column are calculated (steps given on pages 176-177) Choose the row/column, which has the maximum value for allocation. In this case there are five penalties, which have the maximum value 2. The cell with least cost is Row 3 and hence select cell (3,4) for allocation. The supply and demand are 500 and 300 respectively and hence allocate 300 in cell (3,4) as shown in Table 6.22

Table 6.22: Penalty Calculation for each Row and Column

		Destination				Supply	Penalty
		1	2	3	4		
Source	1	4	2	7	3	250	(1)
	2	3	7	5	8	450	(2)
	3	9	4	3	1 300	500	200
Demand		200	400	300	300	0	
		(1)	(2)	(2)	(2)		

Since the demand is satisfied for destination 4, delete column 4 . Now again calculate the penalties for the remaining rows and columns.

Table 6.23: Exhausted Column 4 Deleted

		Destination			Supply	Penalty
		1	2	3		
Source	1	4 250	2 250	7	250	(2)
	2	3	7	5	450	(2)
	3	9	4	3	200	(1)
Demand		200	400	300	150	
		(1)	(2)	(2)		

In the Table 6.24 shown, there are four maximum penalties of values which is 2. Selecting the least cost cell, (1,2) which has the least unit transportation cost 2. The cell (1, 2) is selected for allocation as shown in Table 6.23. Table 6.24 shows the reduced table after deleting row 1.

Table 6.24: Row 1 Deleted

		Destination			Supply	Penalty
Source		1	2	3		
		2	3	7	5	
	2	200			450 250	(2)
	3	9	4	3		(1)
	Demand	200	150	300	200	
		(6) ↑	(3)	(2)		

After deleting column 1 we get the table as shown in the Table 6.25 below.

Table 6.25: Column 1 Deleted

		Destination		Supply	Penalty
Source		2	3		
		2	7	5	
	2			250	(2)
	3	4	3	200	
		150		50	(1)
	Demand	150	300		
		(3) ↑	(2)		

Finally we get the reduced table as shown in Table 6.26

Table 6.26: Final Reduced Table

		Destination		Supply	
Source		2	3		
		2	5		
	2			250	
	3	3		50	
		50		0	
	Demand	300		0	

The initial basic feasible solution is shown in Table 6.27.

Table 6.27: Initial Basic Feasible Solution

Destination					Supply
	W ₁	W ₂	W ₃	W ₄	
F ₁	140				140 (4) (4) (8) (48) (48)
17	5		9	65	
F ₂	50				210
20	10		12	65	260 (2) (2) (8) (45) (45)
F ₃	10	100		250	
15	0	5		65	360 (5) (5) (10) (50) __
F ₄	220				220 (9) __ __ __ __
13	1	10		65	
Demand	200	320	250	210	
	(2)	(1)	(4)	(0)	
	(2)	(5)	(4)	(0)	
	(2)	—	—	(0)	
	(2)	—	—	(0)	
	(3)	—	—	(0)	

$$\begin{aligned}
 \text{Transportation cost} &= (2 \times 250) + (3 \times 200) + (5 \times 250) + (4 \times 150) + (3 \times 50) + \\
 &\quad (1 \times 300) \\
 &= 500 + 600 + 1250 + 600 + 150 + 300 \\
 &= \text{Rs. } 3,400.00
 \end{aligned}$$

Example 5: Find the initial basic solution for the transportation problem and hence solve it.

Table 6.28: Transportation Problem

Destination					Supply	
	1	2	3	4		
Source	1	4	2	7	3	250
	2	3	7	5	8	450
	3	9	4	3	1	500
Demand	200	400	300	300		

Solution: Vogel's Approximation Method (VAM) is preferred to find initial feasible solution. The advantage of this method is that it gives an initial solution which is nearer to an optimal solution or the optimal solution itself.

Step 1: The given transportation problem is a balanced one as the sum of supply equals to sum of demand.

Step 2: The initial basic solution is found by applying the Vogel's Approximation method and the result is shown in Table 6.29.

Table 6.29: Initial Basic Solution Found by Applying VAM

Transportation Model

		Destination				Supply
		1	2	3	4	
Source	1	4 250	2	7		3
	2	3 200	7 250	5 250		450
	3	9 150	4 150	3 50	1 300	500
		Demand	200	400	300	300

Step 3: Calculate the Total Transportation Cost.

$$\begin{aligned}
 \text{Initial Transportation cost} &= (2 \times 250) + (3 \times 200) + (5 \times 250) + (4 \times 150) + \\
 &\quad (3 \times 50) + (1 \times 300) \\
 &= 500 + 600 + 1250 + 600 + 150 + 300 \\
 &= \text{Rs. 3,400}
 \end{aligned}$$

Step 4: Check for degeneracy. For this, verify the condition,

$$\text{Number of allocations, } N = m + n - 1$$

$$6 = 3 + 4 - 1$$

$$6 = 6$$

Since the condition is satisfied, degeneracy does not exist.

Step 5: Test for optimality using modified distribution method. Compute the values of U_i and V_j for rows and columns respectively by applying the formula for occupied cells.

$$C_{ij} + U_i + V_j = 0$$

Then, the opportunity cost for each unoccupied cell is calculated using the formula $\bar{C}_{ij} = C_{ij} + U_i + V_j$ and denoted at the left hand bottom corner of each unoccupied cell. The computed values of u_j and v_i are shown in Table 6.30.

Table 6.30: Calculation of the Opportunity Cost

		Destination				Supply
		1	2	3	4	
Source	1	4 250	2	7	3	250 $U_1 = 2$
	2	3 200	7 250	5 250	8 5	450 $U_2 = -2$
	3	9 150	4 150	3 50	1 300	500 $U_3 = 0$
		Demand	200	400	300	300
			$V_1 = 1$	$V_2 = 4$	$V_3 = -3$	$V_4 = 1$

Calculate the values of U_i and V_j , using the formula for occupied cells. Assume any one of U_i and V_j value as zero (U_3 is taken as 0)

$$\begin{aligned}
 C_{ij} + U_i + V_j &= 0 \\
 4 + 0 + V_2 &= 0, V_2 = -4 \\
 5 + V_2 - 3 &= 0, U_2 = -2 \\
 3 - 2 + V_1 &= 0, V_1 = -1 \\
 2 - 4 + U_1 &= 0, U_1 = 2
 \end{aligned}$$

Calculate the values of \bar{C}_{ij} , using the formula for unoccupied cells

$$\begin{aligned}
 \bar{C}_{ij} &= C_{ij} + U_i + V_j \\
 C_{11} &= 4+2-1=5 \\
 C_{13} &= 7+2-3=6 \\
 C_{14} &= 3+2-1=4 \\
 C_{22} &= 7-2-4=1 \\
 C_{24} &= 8-2-1=5 \\
 C_{31} &= 9+0-1=8
 \end{aligned}$$

Since all the opportunity cost, \bar{C}_{ij} values are positive the solution is optimum.

$$\begin{aligned}
 \text{Total transportation cost} &= (2 \times 25) + (3 \times 200) + (5 \times 250) + (4 \times 150) + (3 \times 50) \\
 &\quad + (1 \times 300) \\
 &= 50 + 600 + 1250 + 600 + 150 + 300 \\
 &= \text{Rs } 2,950/
 \end{aligned}$$

Example 6: Find the initial basic feasible solution for the transportation problem given in Table 6.31.

Table 6.31: Transportation Problem

From	To			Available
	A	B	C	
I	50	30	220	1
II	90	45	170	3
III	250	200	50	4
Requirement	4	2	2	

Solution : The initial basic feasible solution using VAM is shown in Table 6.32.

Table 6.32: Initial Basic Feasible Solution Using VAM

From	I	To			Available
		A	B	C	
Requirement	I	50	30	220	✓(20) (20)
	II	90	45	170	✓(45) (45)
	III	250	200	50	✓(150) (50)
		40	0	20	0
		(40)	(15)	(120)	
		(40)	(15)	--	

Check for degeneracy,

Transportation Model

The number of allocations, N must be equal to $m + n - 1$.

$$\text{i.e., } N = m+n-1$$

$$5 = 3+3 - 1$$

since $4 \neq 5$, therefore degeneracy exists.

To overcome degeneracy, the condition $N = m + n - 1$ is satisfied by allocating a very small quantity, close to zero in an occupied independent cell. (i.e., it should not form a closed loop) or the cell having the lowest transportation cost. This quantity is denoted by e .

This quantity would not affect the total cost as well as the supply and demand values. Table 6.33 shows the resolved degenerate table.

Table 6.33: Resolved Degenerate Table

		To			Available
		A	B	C	
From	I	50	30	220	1
	II	90	45	170	3
	III	250	200	50	4
Requirement		4	2	2	

$$\begin{aligned}
 \text{Total transportation cost} &= (50 \times 1) + (90 \times 3) + (200 \times 2) + (50 \times 2) + (250 \times e) \\
 &= 50 + 270 + 400 + 100 + 250e \\
 &= 820 + 250e = \text{Rs. 820 since } e \rightarrow 0
 \end{aligned}$$

Example 7: Obtain an optimal solution for the transportation problem by MODI method given in Table 6.34.

Table 6.34: Transportation Problem

		Destination				Supply
		D ₁	D ₂	D ₃	D ₄	
Source	S ₁	19	30	50	10	7
	S ₂	70	30	40	60	9
	S ₃	40	8	70	20	18
Demand		5	8	7	14	

Solution:

Step1: The initial basic feasible solution is found using Vogel's Approximation Method as shown in Table 6.35.

**Table 6.35: Initial Basic Feasible Solution Using VAM
Destination**

		D ₁	D ₂	D ₃	D ₄	Supply	
		S ₁	19	30	50	10	7(9)(9)(40)(40)
Source	S ₂	70	30	40	60	2(10)(20)(20)(20)	
	S ₃	40	8	70	20	18(12)(20)(50) -- 16 0	
	Demand	5 0	8 0	7 0	14	42 0	
		(21)	(22)	(10)	(10)		
		(21)	--	(10)	(10)		
		--	--	(10)	(10)		
		--	--	(10)	(50)↑		

$$\begin{aligned}
 \text{Total transportation cost} &= (19 \times 5) + (10 \times 2) + (40 \times 7) + (60 \times 2) + (8 \times 8) + (20 \times 10) \\
 &= 95 + 20 + 280 + 120 + 64 + 200 \\
 &= \text{Rs. 779.00}
 \end{aligned}$$

Step 2: To check for degeneracy, verify the number of allocations, $N = m+n - 1$.
In this problem, number of allocation is 6 which is equal $m+n - 1$.

$$\therefore N = m + n - 1$$

$$6 = 3 + 4 - 1$$

6 = 6 therefore degeneracy does not exist.

Step 3: Test for optimality using MODI method. In Table 6.36 the values of U_i and V_j are calculated by applying the formula $C_{ij} + U_i + V_j = 0$ for occupied cells, and $\bar{C}_{ij} = C_{ij} + U_i + V_j$ for unoccupied cells respectively.

Table 6.36: Optimality Test Using MODI Method
Destination

		D ₁	D ₂	D ₃	D ₄	Supply	
		S ₁	19	30	50	10	7 U ₁ = 0
Source	S ₂	5	32	60	2		
	S ₃	1	-18	7	2	9U ₂ = -50	
	Demand	5	8	7	14	18 U ₃ = -10	

$$V_1 = -19 \quad V_2 = 2 \quad V_3 = 10 \quad V_4 = -10$$

Find the values of the dual variables U_i and V_j for occupied cells.

Initially assume $U_i = 0$,

$$C_{ij} + U_i + V_j = 0,$$

$$\begin{aligned}
 19 + 0 + V_i &= 0, & V_1 &= -19 \\
 10 + 0 + V_4 &= 0, & V_4 &= -10 \\
 60 + U_2 - 10 &= 0, & U_2 &= -50 \\
 20 + U_3 - 10 &= 0, & U_3 &= -10 \\
 8 - 10 + V_2 &= 0, & V_2 &= 2 \\
 40 - 50 + V_3 &= 0, & V_3 &= 10
 \end{aligned}$$

Find the values of the opportunity cost, \bar{C}_{ij} for unoccupied cells,

$$\begin{aligned}
 \bar{C}_{ij} &= C_{ij} + U_i + V_j \\
 C_{12} &= 30 + 0 + 2 = 32 \\
 C_{13} &= 50 + 0 + 10 = 60 \\
 C_{21} &= 70 - 50 - 19 = 1 \\
 C_{22} &= 30 - 50 + 2 = -18 \\
 C_{31} &= 40 - 10 - 19 = 11 \\
 C_{33} &= 70 - 10 + 10 = 70
 \end{aligned}$$

In Table the cell (2,2) has the most negative opportunity cost. This negative cost has to be converted to a positive cost without altering the supply and demand value.

Step 4: Construct a closed loop . Introduce a quantity $+q$ in the most negative cell (S_2, D_2) and a put $-q$ in cell (S_3, D_2) in order to balance the column D_2 . Now, take a right angle turn and locate an occupied cell in column D_4 . The occupied cell is (S_3, D_4) and put $+q$ in that cell. Now, put $-q$ in cell (S_2, D_4) to balance the column D_4 . Join all the cells to have a complete closed path. The closed path is shown in Figure 6.5.

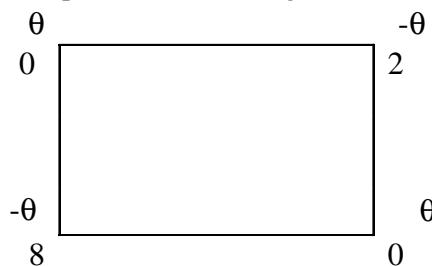


Figure 6.5: Closed Path

Now, identify the $-q$ values, which are 2 and 8. Take the minimum value, 2 which is the allocating value. This value is then added to cells (S_2, D_2) and (S_3, D_4) which have '+' signs and subtract from cells (S_2, D_4) and (S_3, D_2) which have '-' signs. The process is shown in Figure 6.6

(S_2, D_2)	$0+2 = 2$	(S_2, D_4)	$2-2 = 0$
(S_3, D_2)	$8-2 = 6$	(S_3, D_4)	$10+2 = 12$

Figure 6.6

Table 6.37: Closed Path

		Destination				
		D ₁	D ₂	D ₃	D ₄	Supply
Source	S ₁	19 5	30 32	50 60	10 2	7
	S ₂	70 1	30 θ	40 7	60 2 - θ	9
S ₃	40 11	8 -θ	70 8	20 10	18	18
		Demand	5	8	7	14

The table after reallocation is shown in Table 6.38

Table 6.38: After Reallocation

		Destination				
		D ₁	D ₂	D ₃	D ₄	Supply
Source	S ₁	19 5	30	50	10	7
	S ₂	70	30	40	60	9
S ₃	40	8	70	20	12	18
		Demand	5	8	7	14

Now, again check for degeneracy. Here allocation number is 6.

Verify whether number of allocations,

$$N = m + n - 1$$

$$6 = 3 + 4 - 1$$

$$6 = 6$$

therefore degeneracy does not exists.

Again find the values of U_i , V_j and \bar{C}_{ij} for the Table 6.39 shown earlier.

For occupied cells, $C_{ij} + U_i + V_j = 0$

$$19 + 0 + V_1 = 0, \quad V_1 = -19$$

$$10 + 0 + V_4 = 0, \quad V_4 = -10$$

$$20 + U_3 - 10 = 0, \quad U_3 = -10$$

$$8 - 10 + V_2 = 0, \quad V_2 = 2$$

$$30 + U_2 + 2 = 0, \quad U_2 = -32$$

$$40 - 50 + V_3 = 0, \quad V_3 = -10$$

For unoccupied cells, $\bar{C}_{ij} = C_{ij} + U_i + V_j$

$$C_{12} = 30 + 0 + 20 = 50$$

$$C_{13} = 50 + 0 - 8 = 42$$

$$C_{21} = 70 - 32 - 19 = 19$$

$$C_{24} = 60 - 32 - 10 = 18$$

$$C_{31} = 40 - 10 - 19 = 11$$

$$C_{33} = 70 - 10 - 8 = 52$$

The values of the opportunity cost \bar{C}_{ij} are positive. Hence the optimality is reached.

The final allocations are shown in Table 6.39.

Table 6.39: Final Allocation

		Destination				Supply
		D ₁	D ₂	D ₃	D ₄	
Source	S ₁	19 5	30	50	10	7 U ₁ = 0
	S ₂	70 2	30	40 7	60	9 U ₂ = -32
	S ₃	40 6	8	70 12	20	18 U ₃ = -10
Demand		5	8	7	14	

$$V_1 = -19 \quad V_2 = 2 \quad V_3 = -8 \quad V_4 = -10$$

$$\begin{aligned} \text{Total transportation cost} &= (19 \times 5) + (10 \times 2) + (30 \times 2) + (40 \times 7) + (8 \times 6) \\ &\quad + (20 \times 12) \\ &= 95 + 20 + 60 + 280 + 48 + 240 \\ &= \text{Rs. 743} \end{aligned}$$

Example 8: Solve the transportation problem

		Destination			Supply
		1	2	3	
Source	1	3	5	7	10
	2	11	8	9	8
	3	13	3	9	5
Demand		5	9	11	23 25

The problem is unbalanced if $\sum a_i = \sum b_j$, that is, when the total supply is not equal to the total demand. Convert the unbalanced problem into a balanced one by adding a dummy row or dummy column as required and solve.

Here the supply does not meet the demand and is short of 2 units. To convert it to a balanced transportation problem add a dummy row and assume the unit cost for the dummy cells as zero as shown in Table 6.40 and solve.

Table 6.40: Dummy Row Added to TP

		Destination			Supply
		1	2	3	
Source	1	3	5	7	10
	2	11	8	9	8
	3	13	3	9	5
	4	0	0	0	2
Demand		5	9	11	25
					25

6.12 MAXIMIZATION TRANSPORTATION PROBLEM

In this type of problem, the objective is to maximize the total profit or return. In this case, convert the maximization problem into minimization by subtracting all the unit cost from the highest unit cost given in the table and solve.

Example 9: A manufacturing company has four plants situated at different locations, all producing the same product. The manufacturing cost varies at each plant due to internal and external factors. The size of each plant varies, and hence the production capacities also vary. The cost and capacities at different locations are given in the following table:

Table 6.41: Cost and Capacity of Different Plants

Particulars	Plant			
	A	B	C	D
Production cost per unit (Rs.)	18	17	15	12
Capacity	150	250	100	70

The company has five warehouses. The demands at these warehouses and the transportation costs per unit are given in the Table 6.42 below. The selling price per unit is Rs. 30/-

Table 6.42: Transportation Problem

Warehouse	Transportation cost (Rs) — Unit-wise				Demand
	A	B	C	D	
1	6	9	5	3	100
2	8	10	7	7	200
3	2	6	3	8	120
4	11	6	2	9	80
5	3	4	8	10	70

- (i) Formulate the problem to maximize profits.
- (ii) Determine the solution using TORA.
- (iii) Find the total profit.

- (i) The objective is to maximize the profits. Formulation of transportation problem as profit matrix table is shown in Table 6.43. The profit values are arrived as follows.

Profit = Selling Price – Production cost – Transportation cost

Table 6.43: Profit Matrix

Destination

	A	B	C	D	Demand
1	6	4	10	15	100
2	4	3	8	11	200
3	10	7	12	10	120
4	1	7	13	9	80
5	9	9	7	8	70
Supply	150	250	100	70	570

Converting the profit matrix to an equivalent loss matrix by subtracting all the profit values from the highest value 13. Subtracting all the values from 13, the loss matrix obtained is shown in the Table 6.44

Table 6.44: Loss Matrix

Destination

	A	B	C	D	Demand
1	9	11	5	0	100
2	11	12	7	4	200
3	5	8	3	5	120
4	14	8	2	6	80
5	6	6	8	7	70
Supply	150	250	100	70	570

- (ii) To determine the initial solution using TORA

Input Screen:

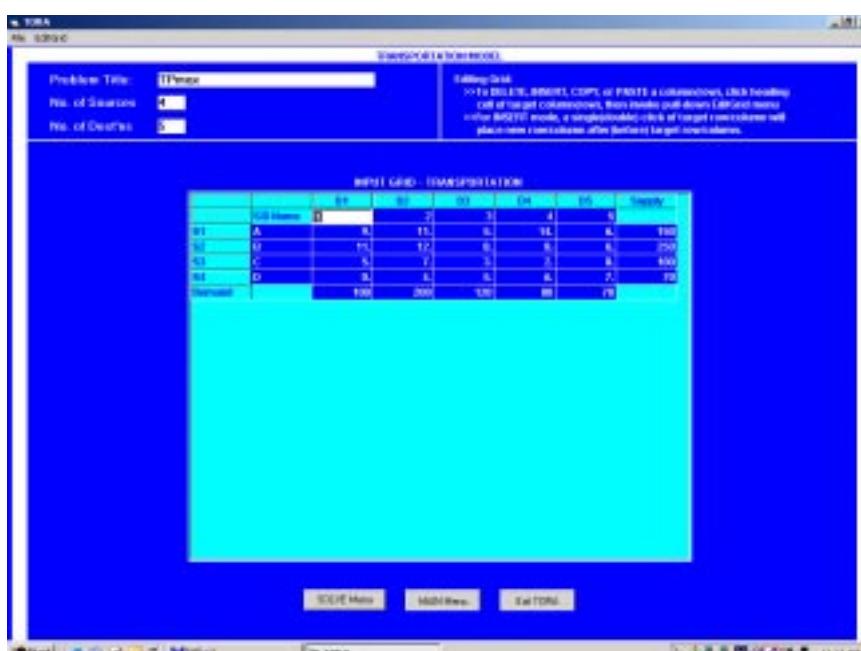


Figure 6.7: TORA, Input Screen for TP Max Problem

Output Screen:



Figure 6.8: TORA Output Screen (Vogel's Method)

The first iteration itself is optimal, hence optimality is reached.

- (iii) To find the total cost:

The total maximization profit associated with the solution is

$$\begin{aligned}
 \text{Total Profit} &= (6 \times 10) + (4 \times 20) + (10 \times 120) + (3 \times 180) + (9 \times 70) + (10 \times 20) \\
 &\quad + (13 \times 80) + (15 \times 70) \\
 &= 60 + 80 + 1200 + 540 + 630 + 200 + 1040 + 1050 \\
 &= \text{Rs } 4800.00
 \end{aligned}$$

6.13 PROHIBITED ROUTES PROBLEM

In practice, there may be routes that are unavailable to transport units from one source to one or more destinations. The problem is said to have an unacceptable or prohibited route. To overcome such kind of transportation problems, assign a very high cost to prohibited routes, thus preventing them from being used in the optimal solution regarding allocation of units.

Example 10: The following transportation table shows the transportation cost per unit (in Rs.) from sources 1, 2, and 3 to destinations A, B, C. Shipment of goods is prohibited from source 2 to destination C. Solve the transportation problem using TORA

Table 6.45: Problem for TORA Solution

Destination

	A	B	C	Supply
1	25	21	19	120
2	15	7		150
3	10	12	16	80
Demand	150	125	75	

Solution: The entries of the transportation cost are made using TORA

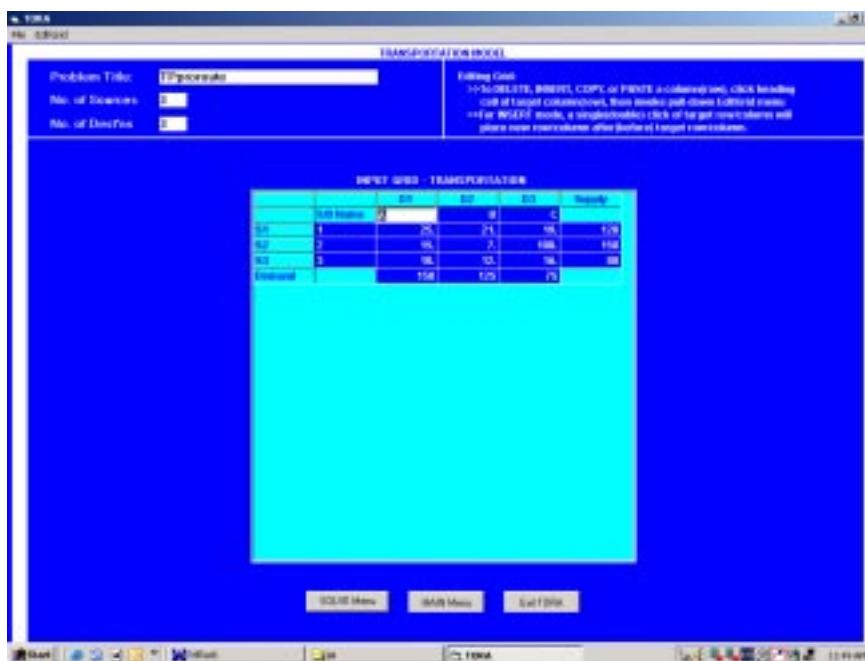


Figure 6.9: TP Prohibited Route TORA (Input Screen)

Output Screen:

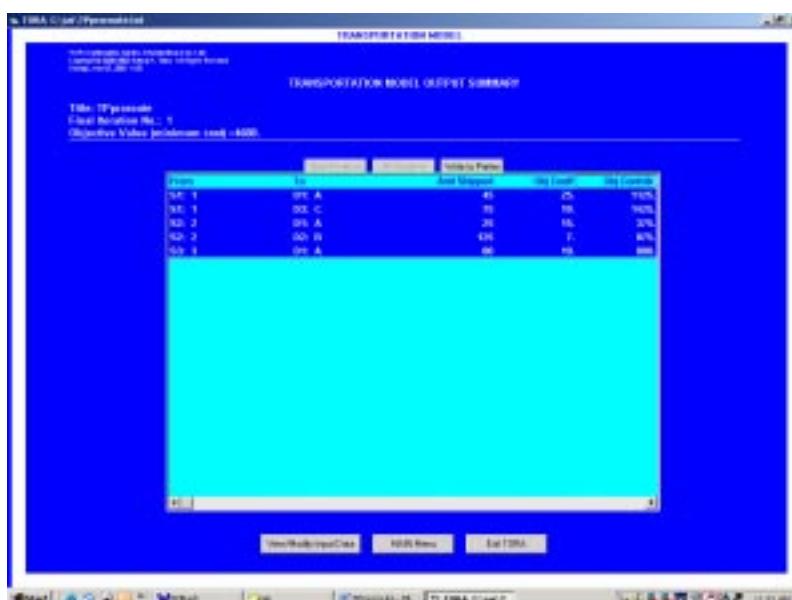


Figure 6.10: TP Prohibited Route (TORA Output Screen)

From the output Schedule, there are no goods that are to be shipped from source 2 to destination C. The total transportation cost is Rs 4600/-

6.14 TRANSHIPMENT PROBLEM

The transshipment problem is an extension of the transportation problem in which the commodity can be transported to a particular destination through one or more intermediate or transshipment nodes.

Each of these nodes in turn supply to other destinations. The objective of the transshipment problem is to determine how many units should be shipped over each node so that all the demand requirements are met with the minimum transportation cost.

Considering a company with its manufacturing facilities situated at two places, Coimbatore and Pune. The units produced at each facility are shipped to either of the company's warehouse hubs located at Chennai and Mumbai. The company has its own retail outlets in Delhi, Hyderabad, Bangalore and Thiruvananthapuram. The network diagram representing the nodes and transportation per unit cost is shown in Figure 6.11. The supply and demand requirements are also given.

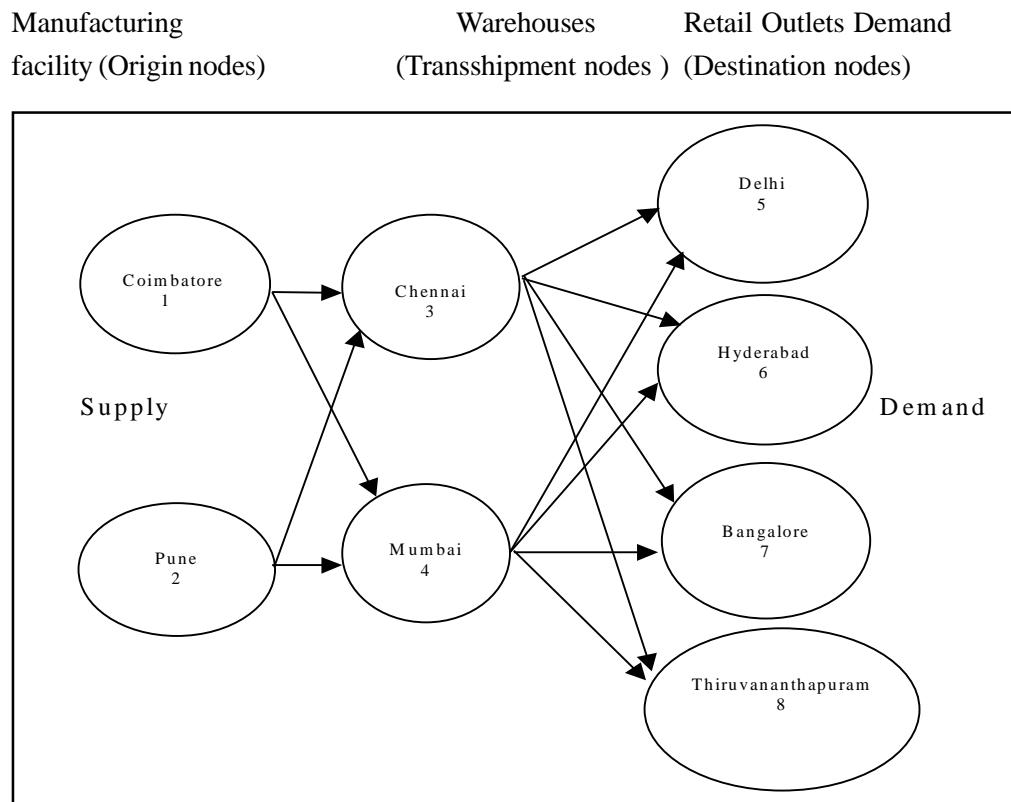


Figure 6.11: Network Representation of Transshipment Problem

Solving Transshipment Problem using Linear Programming

Let

X_{ij} be the number of units shipped from node i to node j,

X_{13} be the number of units shipped from Coimbatore to Chennai,

X_{24} be the number of units shipped from Pune to Mumbai, and so on

Table 6.46 shows the unit transportation cost from sources to destination.

Table 6.46: TP of the Shipment

	Warehouse	
Facility	Chennai	Mumbai
Coimbatore	4	7
Pune	6	3

Warehouses	Retail outlets			
	Delhi	Hyderabad	Bangalore	Thiruvananthapuram
Chennai	7	4	3	5
Mumbai	5	6	7	8

Objective

The objective is to minimize the total cost

Minimize

$$Z = 4X_{13} + 7X_{14} + 6X_{23} + 3X_{24} + 7X_{35} + 4X_{36} + 3X_{37} + 5X_{38} + 5X_{45} + 6X_{46} + 7X_{47} + 8X_{48}$$

Constraints: The number of units shipped from Coimbatore must be less than or equal to 800. Because the supply from Coimbatore facility is 800 units. Therefore, the constraints equation is as follows:

$$X_{13} + X_{14} \leq 800 \quad \dots \dots \dots \text{(i)}$$

Similarly, for Pune facility

$$X_{23} + X_{24} \leq 600 \quad \dots \dots \dots \text{(ii)}$$

Now, considering the node 3,

Number of units shipped out from node 1 and 2 are,

$$X_{13} + X_{23}$$

Number of units shipped out from node 3 is,

$$X_{35} + X_{36} + X_{37} + X_{38}$$

The number of units shipped in must be equal to number of units shipped out, therefore

$$X_{13} + X_{23} = X_{35} + X_{36} + X_{37} + X_{38}$$

Bringing all the variables to one side, we get

$$-X_{13} - X_{23} + X_{35} + X_{36} + X_{37} + X_{38} = 0 \quad \dots \dots \dots \text{(iii)}$$

Similarly for node 4

$$-X_{14} - X_{24} + X_{45} + X_{46} + X_{47} + X_{48} = 0 \quad \dots \dots \dots \text{(iv)}$$

Now considering the retail outlet nodes, the demand requirements of each outlet must be satisfied. Therefore for retail node 5, the constraint equation is

$$X_{35} + X_{45} = 350 \quad \dots \dots \dots \text{(v)}$$

Similarly for nodes 6, 7, and 8, we get,

$$X_{36} + X_{46} = 200 \quad \dots \dots \dots \text{(vi)}$$

$$X_{37} + X_{47} = 400 \quad \dots \dots \dots \text{(vii)}$$

$$X_{38} + X_{48} = 450 \quad \dots \dots \dots \text{(viii)}$$

Linear Programming formulation,

$$\text{Minimize } Z = 4X_{13} + 7X_{14} + 6X_{23} + 3X_{24} + 7X_{35} + 4X_{36} + 3X_{37} + 5X_{38} + 5X_{45} + 6X_{46} + 7X_{47} + 8X_{48}$$

Subject to constraints ,

$$\begin{aligned} X_{13} + X_{14} &\leq 800 \\ X_{23} + X_{24} &\leq 600 \end{aligned} \quad \left. \right\} \text{origin constraints}$$

$$-X_{13} - X_{23} + X_{35} + X_{36} + X_{37} + X_{38} = 0$$

$$-X_{14} - X_{24} + X_{45} + X_{46} + X_{47} + X_{48} = 0$$

$$X_{35} + X_{45} = 350$$

$$X_{36} + X_{46} = 200$$

$$X_{37} + X_{47} = 400$$

$$X_{38} + X_{48} = 450$$

Check Your Progress 6.2

1. In the transportation model an example of decision under certainty or decision-making under uncertainty.
2. How can the travelling sales man problem be solved using transportation model.

Notes: (a) Write your answer in the space given below.
(b) Please go through the lesson sub-head thoroughly you will get your answers in it.
(c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

6.15 LET US SUM UP

The use of transportation models to minimize the cost of shipping from a number of source to number of destinations. In most general form, a transportation problems has a number of origins and a number of destination. A number of techniques are there to compute the initial basic feasible solution of a TP. These are NWC, LCM, VAM. Further there can be an optimum solution while could obtained from MODI and stepping stone. Transportation problem can be generalized into transshipment problem where shipment could be feasible from origin to origin.

6.16 LESSON-END ACTIVITY

I hope you all are familiar with the Aeroplane and Airport. The Airport authorities take lot of pain in streamlining and maintaining traffic. So that the havoc situations could be controlled and also there may not be any confusion among each other. Being an expert in Transportation know this. Transportation programming techniques facilitates. in maintaining the traffic rules. Apply with the help of illustrations.

6.17 KEYWORDS

- Origin** : Origin of a TP is the from which shipments are dispatched.
- Destination** : Destination of TA is a point to where shipment are transported.
- Source** : Supply location is a TP.
- Northwest corner** : A systematic procedure for establishing our initial feasible solution to an optimal
- Degeneracy** : A situation that occurs where the number of occupied squares in any solution is less than number of row plus number of column in a transportation basic.

Unbalance problem : A situation is which demand is not equal to supply.

Transportation Model

Summary Destination : An artificial destination.

VAM : Vogel Approximation Method is an interactive procedure of a feasible solution.

6.18 QUESTIONS FOR DISCUSSION

1. **Explain the initial basic feasible solution of transportation model.**
2. **Is the TP model is an example of decision-making under certainty or decision-making under uncertainty? Why?**
3. **Write True or False against each statement:**
 - (a) TP is a special type of linear programming
 - (b) Dummy rows to dummy column are assigned source values.
 - (c) Initial Basic solution can be obtained by MODI method.
 - (d) Least cost method is a best method to find initial basic solution.
 - (e) In maximisation the objective is to maximise the total profit.
4. **Briefly comment on the following statements:**
 - (a) Transportation problem is said to be unbalanced.
 - (b) Optimum solution have an edge as compared to initial basic feasible solution.
 - (c) Transportation problem can be generalized with a transshipment problem.
 - (d) Problem is known as unbalanced TP if they are unequal.
 - (e) MODI distribution method provides a minimum cost solution.
 - (f) Degeneracy does not cause any serious difficulty.
 - (g) Transportation problem is balanced when sum of supply equals to sum of demand.
5. **Fill in the blanks:**
 - (a) In the transportation problem _____ are always transported
 - (b) Initial basis feasible solution through VAM will be _____
 - (c) Demand variation may occur because of change in _____ preference
 - (d) TP deals with the transportation of a _____ manufactured.
 - (e) In real life supply & demand requirement will be rarely _____
6. **Differentiate between the following:**
 - (a) MODI vs Shipping Stone
 - (b) LCM vs NWC
 - (c) VAM vs MODI

6.19 TERMINAL QUESTIONS

1. What is a transportation problem ?
2. What is the difference between a balanced transportation problem and an unbalanced transportation problem ?

3. What are the methods used to find the initial transportation cost ?
4. Which of the initial three methods give a near optimal solution ?
5. Explain Vogel's approximation method of finding the initial solution.
6. What is degeneracy in a transportation problem ? How is it resolved ?
7. What are the conditions for forming a closed loop ?
8. How are the maximization problems solved using transportation model ?
9. How is optimality tested in solving transportation problems ?
10. In what ways is a transshipment problem different from a transportation problem ?

Exercise Problems

1. Develop a network representation of the transportation problem for a company that manufactures products at three plants and ships them to three warehouses. The plant capacities and warehouse demands are shown in the following table:

The transportation cost per unit (in Rs.) is given in matrix.

Plant	Warehouse			Plant Capacity (no. of units)
	W1	W2	W3	
P1	22	18	26	350
P2	12	12	10	450
P3	14	20	10	200
Warehouse demand (no. of units)	250	450	300	

2. Determine whether a dummy source or a dummy destination is required to balance the model given.
 - (a) Supply $a_1 = 15, a_2 = 5, a_3 = 4, a_4 = 6$
Demand $b_1 = 4, b_2 = 15, b_3 = 6, b_4 = 10$
 - (b) Supply $a_1 = 27, a_2 = 13, a_3 = 10$
Demand $b_1 = 30, b_2 = 10, b_3 = 6, b_4 = 10$
 - (c) Supply $a_1 = 2, a_2 = 3, a_3 = 5$
Demand $b_1 = 3, b_2 = 2, b_3 = 2, b_4 = 2, b_5 = 1$.
3. A state has three power plants with generating capacities of 30, 40 and 25 million KWH that supply electricity to three cities located in the same state. The demand requirements (maximum) of the three cities are 35, 40 and 20 million KWH. The distribution cost (Rs. in thousand) per million unit for the three cities are given in the table below:

Plant	City		
	1	2	3
1	60	75	45
2	35	35	40
3	55	50	45

(a) Formulate the problem as a transportation model.

Transportation Model

- (b) Determine an economical distribution plan.

- (c) If the demand is estimated to increase by 15%, what is your revised plan?

- (d) If the transmission loss of 5% is considered, determine the optimal plan.

4. Find the initial transportation cost for the transportation matrix given using North-West corner method, Least cost method and Vogel's Approximation Method.

		Destination				
		1	2	3	4	Supply
Source	A	5	6	7	8	25
	B	7	5	4	2	75
	C	6	1	3	2	15
Demand		50	30	20	15	

5. In problem No. 4, if the demand for destination 4 increases from 15 units to 25 units, develop the transportation schedule incorporating the change.
6. Find the initial solution using all the three methods and hence find the optimal solution using TORA package for the following transportation problem. The unit transportation cost is given in the following matrix:

		Warehouse						
		1	2	3	4	5	6	Supply
Factory	A	10	25	35	16	18	22	70
	B	11	22	16	18	22	19	60
	C	21	32	41	20	20	11	50
	D	25	24	23	22	23	24	85
	E	16	21	18	20	19	16	45
Demand		55	45	35	40	70	65	

7. The Sharp Manufacturing Company produces three types of monoblock pumps for domestic use. Five machines are used for manufacturing the pumps. The production rate varies for each machine and also the unit product cost. Daily demand and machine availability are given below.

Demand Information

Product				
	A	B	C	
Demand (units)	2000	15000	700	

Machine Availability Details

Machine capacity (units)					
	1	2	3	4	5
Available	700	1000	1500	1200	800

Unit Product Cost

Machine	Product		
	A	B	C
1	150	80	75
2	120	95	60
3	112	100	60
4	121	95	50
5	125	75	50

Determine the minimum production schedule for the products and machines.

8. A company has plants at locations A, B and C with the daily capacity to produce chemicals to a maximum of 3000 kg, 1000 kg and 2000 kg respectively. The cost of production (per kg) are Rs. 800 Rs. 900 and Rs. 7.50 respectively. Customer's requirement of chemicals per day is as follows:

Customer	Chemical Required	Price offered
1	2000	200
2	1000	215
3	2500	225
4	1000	200

Transportation cost (in rupees) per kg from plant locations to customer's place is given in table.

		Customer			
		1	2	3	4
Plant	A	5	7	10	12
	B	7	3	4	2
	C	4	6	3	9

Find the transportation schedule that minimizes the total transportation cost.

9. A transportation model has four supplies and five destinations. The following table shows the cost of shipping one unit from a particular supply to a particular destination.

Source	Destination					Supply
	1	2	3	4	5	
1	13	6	9	6	10	13
2	8	2	7	7	9	15
3	2	12	5	8	7	13
Demand	10	15	7	10	2	

The following feasible transportation pattern is proposed:

$$x_{11} = 10, x_{12} = 3, x_{22} = 9, x_{23} = 6, x_{33} = 9, x_{34} = 4, x_{44} = 9, x_{45} = 5.$$

Test whether these allocations involve least transportation cost. If not, determine the optimal solution.

10. A linear programming model is given:

Minimize $Z = 8x_{11} + 12x_{12} + 9x_{22} + 10x_{23} + 7x_{31} + 6x_{32} + 15x_{33}$, subject to the constraints,

$$\left. \begin{array}{l} x_{11} + x_{12} + x_{13} = 60 \\ x_{21} + x_{22} + x_{23} = 50 \\ x_{31} + x_{32} + x_{33} = 30 \end{array} \right\} \text{Supply constraints}$$

$$\left. \begin{array}{l} x_{11} + x_{21} + x_{31} = 20 \\ x_{12} + x_{22} + x_{32} = 60 \\ x_{13} + x_{23} + x_{33} = 30 \end{array} \right\} \text{Demand constraints}$$

Formulate and solve as a transportation problem to minimize the transportation cost.

11. A company has four factories situated in four different locations in the state and four company showrooms in four other locations outside the state. The per unit sale price, transportation cost and cost of production is given in table below, along with weekly requirement.

Factory	Showrooms				Cost of production (Rs)
	1	2	3	4	
A	9	4	5	3	12
B	4	4	4	4	17
C	4	6	5	6	19
D	8	7	7	4	17

Factory	Weekly Capacity (units)	Weekly demand (units)
A	15	10
B	20	14
C	25	20
D	20	22

Determine the weekly distribution schedule to maximize the sales profits.

12. Solve the given transportation problem to maximize profit.

Source	Profit / unit						Supply
	1	2	3	4	5	6	
A	65	30	77	31	65	51	200
B	60	51	65	42	64	76	225
C	70	62	21	71	45	52	125
Demand	45	55	40	60	25	70	

Use TORA to solve the problem.

13. A computer manufacturer has decided to launch an advertising campaign on television, magazines and radio. It is estimated that maximum exposure for these media will be 70, 50, and 40 million respectively. According to a market survey, it was found that the minimum desired exposures within age groups 15-20, 21-25, 26-30, 31-35 and above 35 are 10, 20, 25, 35 and 55 million respectively. The table below gives the estimated cost in paise per exposure for each of the media. Determine an advertising plan to minimize the cost.

Media	Age Groups				
	15-20	21-25	26-30	31-35	above 35
TV	14	9	11	11	12
Radio	11	7	6	7	8
Magazine	9	10	7	10	8

Solve the problem and find the optimal solution, i.e., maximum coverage at minimum cost.

14. A garment manufacturer has 4 units I, II, III, and IV, the production from which are received by 4 direct customers. The weekly production of each manufacturing unit is 1200 units and all the units are of the same capacity. The company supplies the entire production from one unit to one supplier. Since the customers are situated at different locations, the transportation cost per unit varies. The unit cost of transportation is given in the table. As per the company's policy, the supply from unit B is restricted to customer 2 and 4, and from unit D to customer 1 and 3. Solve the problem to cope with the supply and demand constraints.

Manufacturing unit	1	2	3	4
A	4	6	8	3
B	4	—	5	—
C	6	5	5	9
D	—	7	—	6

15. Check whether the following transportation problem has an optimal allocation:

Warehouse						
	1	2	3	4	5	Supply
A	100					100
B	25					25
C	25	50				75
D			50	100	50	200
Dummy					100	100
Demand	150	50	50	100	150	

16. A company dealing in home appliances has a sales force of 20 men who operate from three distribution centers. The sales manager feels that 5 salesmen are needed to distribute product line I, 6 to distribute product line II, 5 for product line III and 4 to distribute product line IV. The cost (in Rs) per day of assigning salesmen from each of the offices are as follows:

		Product Line			
		I	II	III	IV
Source	A	10	12	13	9
	B	9	11	12	13
	C	7	8	9	10

Currently, 8 salesmen are available at center A, 5 at center B and 7 at center C. How many salesmen should be assigned from each center to sell each product line, in order to minimize the cost? Is the solution unique?

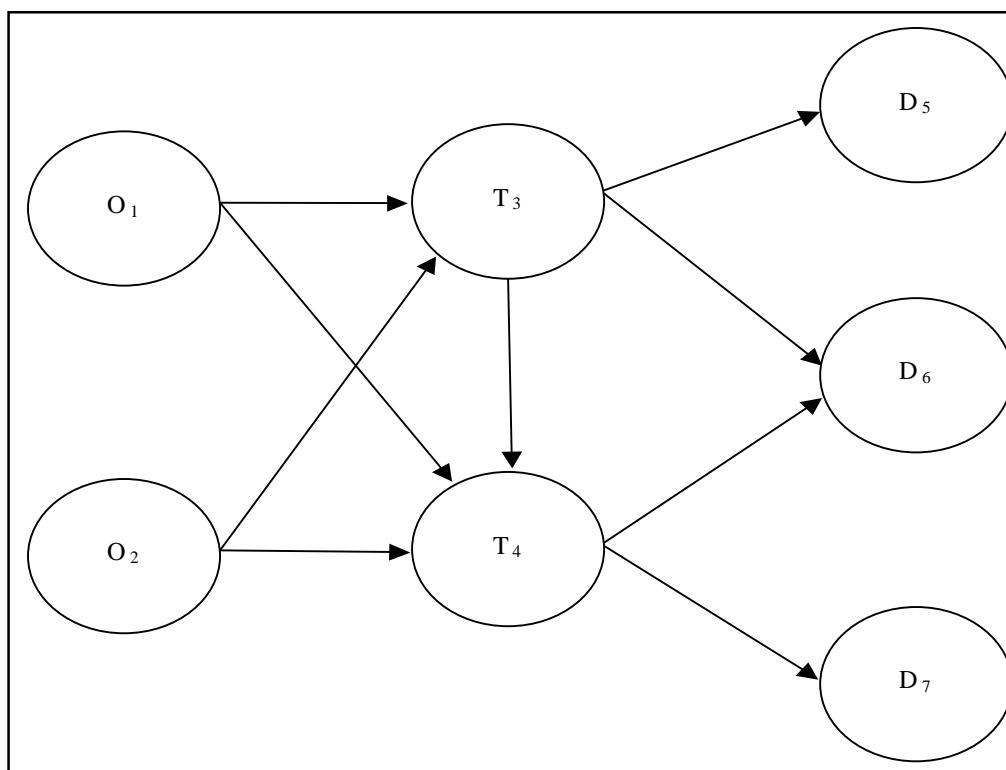
17. Solve the following degenerate transportation problem:

		Destination			
		I	II	III	Supply
Source	A	7	3	4	2
	B	2	1	3	3
		3	4	6	5
Demand		4	1	5	

18. Three water distribution tanks with daily capacities of 7, 6 and 9 lakh litres respectively, supply three distribution areas with daily demands of 5, 8 and 9 lakh litres respectively. Water is transported to the distribution areas through an underground network of pipelines. The cost of transportation is Rs 0.50 per 1000 litres per pipeline kilometer. The table shows the pipeline lengths between the water tanks and the distribution areas.

		Distribution Area		
		1	2	3
Source	A	75	95	120
	B	250	150	80
		300	250	140

- A. Formulate the transportation model
 B. Use TORA to determine the optimum distribution schedule
19. In problem 18, if the demand for distribution area 3 increases to 11 lakh litres, determine a suitable distribution plan to meet the excess demand and minimize the distribution cost. Use TORA to solve the problem.
20. Formulate a linear programming model for the following transshipment network given below.



6.20 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

3. (a) True (b) False (c) False (d) False (e) True
5. (a) consignment (b) least (c) customer (d) product (e) equal

6.21 SUGGESTED READINGS

Render, B. E. Stair, R.M., *Management Science: A self-correcting approach*, Boston Allyn and Bacon, Inc.

Bowman E., *Production Scheduling by the transportation method of LP*, "Operation Research".

Srinivasan V., "A transshipment model for cost management decision." *Management Science*, Vol. 20, June 1974.

Sadleir C.D., *Use of Transportation Method of LP in Production Planning: A Case Study*, "Operation Research" Vol. 21. No. 4.

LESSON

7

ASSIGNMENT MODEL

CONTENTS

- 7.0 Aims and Objectives
- 7.1 Introduction
- 7.2 Mathematical Structure of Assignment Problem
- 7.3 Network Representation of Assignment Problem
- 7.4 Use of Linear Programming to Solve Assignment Problem
- 7.5 Types of Assignment Problem
- 7.6 Hungarian Method for Solving Assignment Problem
- 7.7 Unbalanced Assignment Problem
- 7.8 Restricted Assignment Problem
- 7.9 Multiple and Unique Solutions
- 7.10 Maximization Problem
- 7.11 Travelling Salesman Problem
- 7.12 Solving Problems on the Computer with TORA
- 7.13 Solving Unbalanced Assignment Problem using Computer
- 7.14 Solving Maximization Problems Using Computers
- 7.15 Let us Sum Up
- 7.16 Lesson-end Activity
- 7.17 Keywords
- 7.18 Questions for Discussion
- 7.19 Terminal Questions
- 7.20 Model Answers to Questions for Discussion
- 7.21 Suggested Readings

7.0 AIMS AND OBJECTIVES

In this lesson we would be able to learn assignment of various work activities using various methods of assignment problems. Solving both maximization and minimization problems and both bounded and unbounded solutions of assignment problem.

7.1 INTRODUCTION

The basic objective of an assignment problem is to assign n number of resources to n number of activities so as to minimize the total cost or to maximize the total profit of allocation in such a way that the measure of effectiveness is optimized. The problem of

assignment arises because available resources such as men, machines, etc., have varying degree of efficiency for performing different activities such as job. Therefore cost, profit or time for performing the different activities is different. Hence the problem is, how should the assignments be made so as to optimize (maximize or minimize) the given objective. The assignment model can be applied in many decision-making processes like determining optimum processing time in machine operators and jobs, effectiveness of teachers and subjects, designing of good plant layout, etc. This technique is found suitable for routing travelling salesmen to minimize the total travelling cost, or to maximize the sales.

7.2 MATHEMATICAL STRUCTURE OF ASSIGNMENT PROBLEM

The structure of assignment problem of assigning operators to jobs is shown in Table 7.1.

Table 7.1: Structure of Assignment Problem

		Operator					
		1	2	j	n
Job	1	t_{11}	t_{12}	...	t_{1j}	t_{1n}
	2	t_{21}	t_{22}	t_{2j}	t_{2n}
	.	.					
	I	t_{i1}	t_{i2}		t_{ij}		t_{in}
	.	.					
	n	t_{n1}	t_{n2}		t_{nj}		t_{nn}

Let n be the number of jobs and number of operators.

t_{ij} be the processing time of job i taken by operator j.

A few applications of assignment problem are:

- assignment of employees to machines.
- assignment of operators to jobs.
- effectiveness of teachers and subjects.
- allocation of machines for optimum utilization of space.
- salesmen to different sales areas.
- clerks to various counters.

In all the cases, the objective is to minimize the total time and cost or otherwise maximize the sales and returns.

7.3 NETWORK REPRESENTATION OF ASSIGNMENT PROBLEM

An assignment model is represented by a network diagram in Figure 1 for an operator – job assignment problem, given in Table 7.2 the time taken (in mins.) by operators to perform the job.

Table 7.2: Assignment Problem

Operator	Job		
	1	2	3
A	10	16	7
B	9	17	6
C	6	13	5

The assignment problem is a special case of transportation problem where all sources and demand are equal to 1.

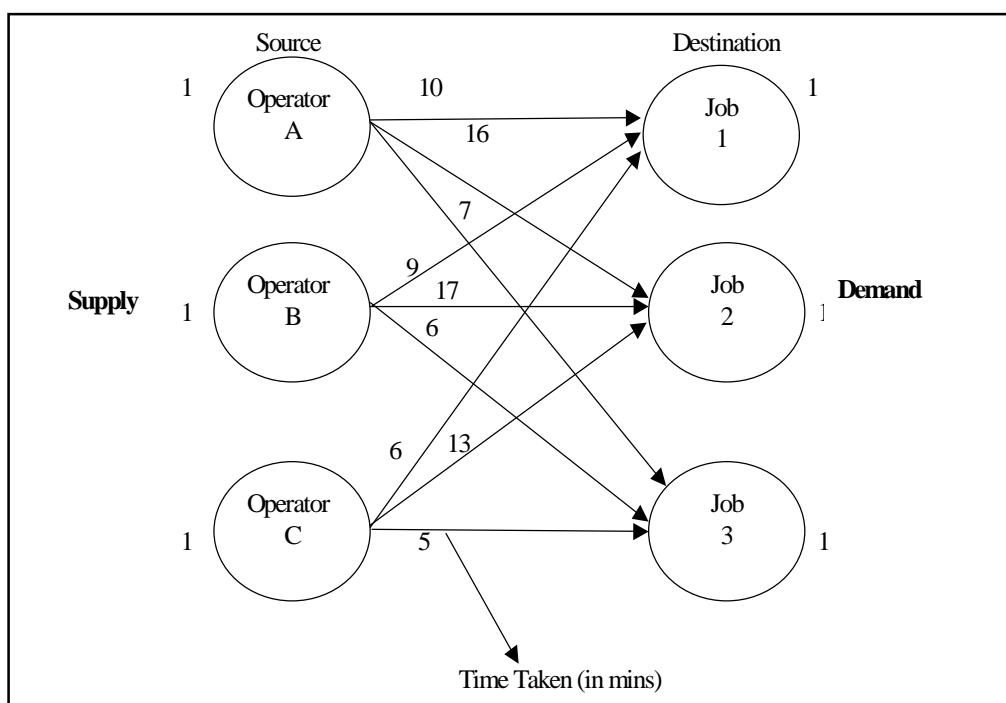


Figure 7.1: Network Diagram for an Operator-job Assignment Problem

7.4 USE OF LINEAR PROGRAMMING TO SOLVE ASSIGNMENT PROBLEM

A linear programming model can be used to solve the assignment problem. Consider the example shown in Table 2, to develop a linear programming model.

Let,

x_{11} represent the assignment of operator A to job 1

x_{12} represent the assignment of operator A to job 2

x_{13} represent the assignment of operator A to job 3

x_{21} represent the assignment of operator B to job 1

and so on.

Formulating the equations for the time taken by each operator,

$$10x_{11} + 16x_{12} + 7x_{13} = \text{time taken by operator A.}$$

$$9x_{21} + 17x_{22} + 6x_{23} = \text{time taken by operator B.}$$

$$6x_{31} + 13x_{32} + 5x_{33} = \text{time taken by operator C.}$$

The constraint in this assignment problem is that each operator must be assigned to only one job and similarly, each job must be performed by only one operator. Taking this constraint into account, the constraint equations are as follows:

$$x_{11} + x_{12} + x_{13} \leq 1 \text{ operator A}$$

$$x_{21} + x_{22} + x_{23} \leq 1 \text{ operator B}$$

$$x_{31} + x_{32} + x_{33} \leq 1 \text{ operator C}$$

$$x_{11} + x_{21} + x_{31} = 1 \text{ Job 1}$$

$$x_{12} + x_{22} + x_{32} = 1 \text{ Job 2}$$

$$x_{13} + x_{23} + x_{33} = 1 \text{ Job 3}$$

Objective function: The objective function is to minimize the time taken to complete all the jobs. Using the cost data table, the following equation can be arrived at:

The objective function is,

$$\text{Minimize } Z = 10 x_{11} + 16 x_{12} + 7 x_{13} + 9 x_{21} + 17 x_{22} + 6 x_{23} + 6 x_{31} + 13 x_{32} + 5 x_{33}$$

The linear programming model for the problem will be,

$$\text{Minimize } Z = 10 x_{11} + 16 x_{12} + 7 x_{13} + 9 x_{21} + 17 x_{22} + 6 x_{23} + 6 x_{31} + 13 x_{32} + 5 x_{33}$$

subject to constraints

$$x_{21} + x_{22} + x_{23} \leq 1 \quad \dots \dots \dots \text{(ii)}$$

$$x_{31} + x_{32} + x_{33} \leq 1 \quad \dots \dots \dots \text{(iii)}$$

where, $x_{ij} \geq 0$ for $i = 1, 2, 3$ and $j = 1, 2, 3$.

The problem is solved on a computer, using transportation model in TORA package. The input screen and output screens are shown in Figure 7.1 and Figure 7.2 respectively.

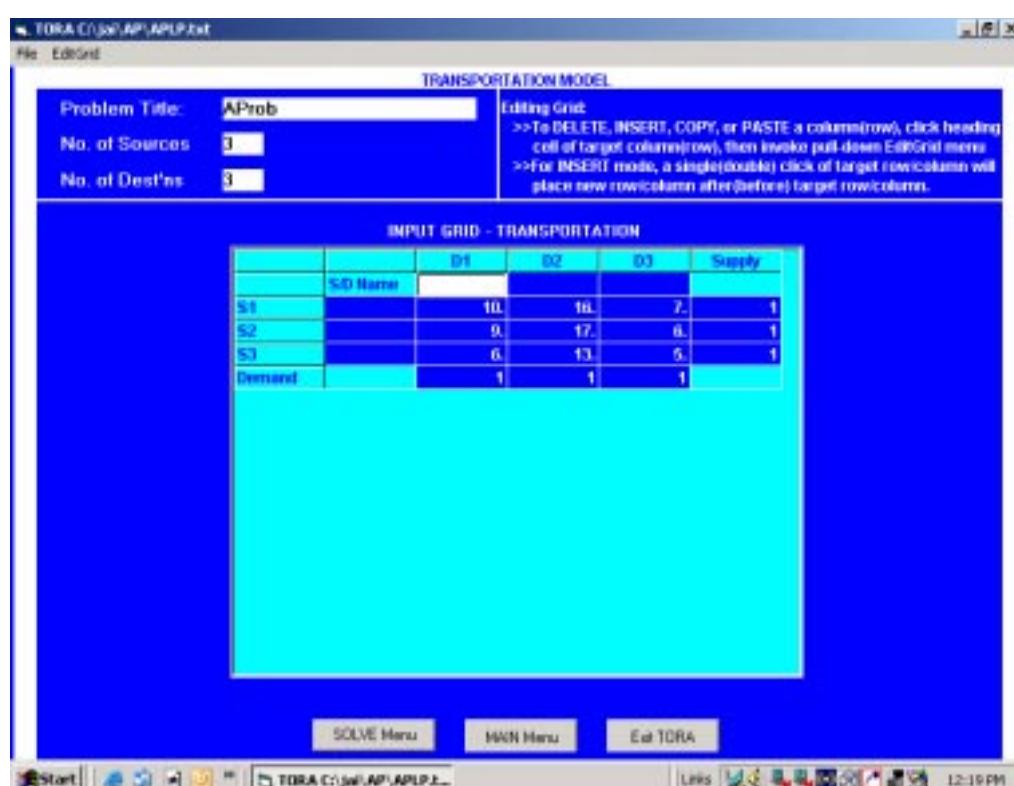
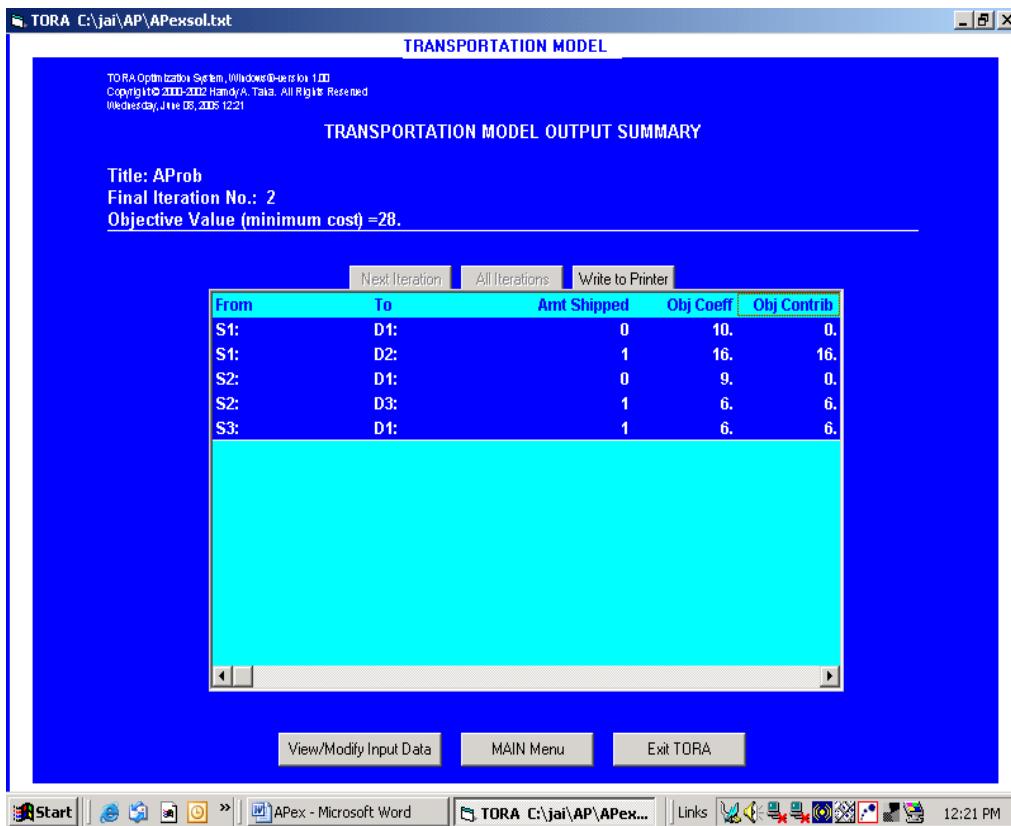


Figure 7.2: TORA, Input Screen

**Figure 7.3: TORA, Output Screen**

The objective function value = 28 mins.

Table 7.3: The Assignment Schedule

Men	Job	Time Taken (in mins.)
1	2	16
2	3	6
3	1	6
	Total	28

7.5 TYPES OF ASSIGNMENT PROBLEM

The assignment problems are of two types (i) balanced and (ii) unbalanced. If the number of rows is equal to the number of columns or if the given problem is a square matrix, the problem is termed as a ***balanced assignment problem***. If the given problem is not a square matrix, the problem is termed as an ***unbalanced assignment problem***.

If the problem is an unbalanced one, add dummy rows /dummy columns as required so that the matrix becomes a square matrix or a balanced one. The cost or time values for the dummy cells are assumed as zero.

7.6 HUNGARIAN METHOD FOR SOLVING ASSIGNMENT PROBLEM

Step 1: In a given problem, if the number of rows is not equal to the number of columns and vice versa, then add a dummy row or a dummy column. The assignment costs for dummy cells are always assigned as zero.

- Step 2:* Reduce the matrix by selecting the smallest element in each row and subtract with other elements in that row.
- Step 3:* Reduce the new matrix column-wise using the same method as given in step 2.
- Step 4:* Draw minimum number of lines to cover all zeros.
- Step 5:* If Number of lines drawn = order of matrix, then optimally is reached, so proceed to step 7. If optimally is not reached, then go to step 6.
- Step 6:* Select the smallest element of the whole matrix, which is **NOT COVERED** by lines. Subtract this smallest element with all other remaining elements that are **NOT COVERED** by lines and add the element at the intersection of lines. Leave the elements covered by single line as it is. Now go to step 4.
- Step 7:* Take any row or column which has a single zero and assign by squaring it. Strike off the remaining zeros, if any, in that row and column (X). Repeat the process until all the assignments have been made.
- Step 8:* Write down the assignment results and find the minimum cost/time.

Note: While assigning, if there is no single zero exists in the row or column, choose any one zero and assign it. Strike off the remaining zeros in that column or row, and repeat the same for other assignments also. If there is no single zero allocation, it means multiple number of solutions exist. But the cost will remain the same for different sets of allocations.

Example 1: Assign the four tasks to four operators. The assigning costs are given in Table 7.4.

Table 7.4: Assignment Problem

		Operators			
		1	2	3	4
Tasks	A	20	28	19	13
	B	15	30	31	28
	C	40	21	20	17
	D	21	28	26	12

Solution:

- Step 1:* The given matrix is a square matrix and it is not necessary to add a dummy row/column
- Step 2:* Reduce the matrix by selecting the smallest value in each row and subtracting from other values in that corresponding row. In row A, the smallest value is 13, row B is 15, row C is 17 and row D is 12. The row wise reduced matrix is shown in Table 7.5.

Table 7.5: Row-wise Reduction

		Operators			
		1	2	3	4
Tasks	A	7	15	6	0
	B	0	15	16	13
	C	23	4	3	0
	D	9	16	14	0

Step 3: Reduce the new matrix given in Table 6 by selecting the smallest value in each column and subtract from other values in that corresponding column. In column 1, the smallest value is 0, column 2 is 4, column 3 is 3 and column 4 is 0. The column-wise reduction matrix is shown in Table 7.6.

Table 7.6: Column-wise Reduction Matrix

		Operators			
		1	2	3	4
Tasks	A	7	11	3	6
	B	0	11	13	13
	C	23	0	0	0
	D	9	12	11	0

Step 4: Draw minimum number of lines possible to cover all the zeros in the matrix given in Table 7.7

Table 7.7: Matrix with all Zeros Covered

		Operators			
		1	2	3	4
Tasks	A	7	11	3	0
	B	0	11	13	13
	C	23	0	0	0
	D	9	12	11	0

No. of lines drawn \neq order of matrix

The first line is drawn crossing row C covering three zeros, second line is drawn crossing column 4 covering two zeros and third line is drawn crossing column 1 (or row B) covering a single zero.

Step 5: Check whether number of lines drawn is equal to the order of the matrix, i.e., $3 \neq 4$. Therefore optimally is not reached. Go to step 6.

Step 6: Take the smallest element of the matrix that is not covered by single line, which is 3. Subtract 3 from all other values that are not covered and add 3 at the intersection of lines. Leave the values which are covered by single line. Table 7.8 shows the details.

Table 7.8: Subtracted or Added to Uncovered Values and Intersection Lines Respectively

		Operators			
		1	2	3	4
Tasks	A	7	9	0	0
	B	0	9	10	13
	C	26	0	0	3
	D	9	9	8	0

Step 7: Now, draw minimum number of lines to cover all the zeros and check for optimality. Here in Table 7.9 minimum number of lines drawn is 4 which is equal to the order of matrix. Hence optimality is reached.

Table 7.9: Optimality Matrix

		Operators			
		1	2	3	4
Tasks	A	7	9	0	0
	B	0	9	10	13
	C	26	0	0	3
	D	9	9	8	0

No. of lines
drawn = order of matrix

Step 8: Assign the tasks to the operators. Select a row that has a single zero and assign by squaring it. Strike off remaining zeros if any in that row or column. Repeat the assignment for other tasks. The final assignment is shown in Table 7.10.

Table 7.10: Final Assignment

		Operators			
		1	2	3	4
Tasks	A	7	9	0	X
	B	0	9	10	13
	C	26	0	X	3
	D	9	9	8	0

Therefore, optimal assignment is:

Task	Operator	Cost
A	3	19
B	1	15
C	2	21
D	4	12
Total Cost = Rs. 67.00		

Example 2: Solve the following assignment problem shown in Table 7.11 using Hungarian method. The matrix entries are processing time of each man in hours.

Table 7.11: Assignment Problem

Assignment Model

		Men				
		1	2	3	4	5
Job	I	20	15	18	20	25
	II	18	20	12	14	15
	III	21	23	25	27	25
	IV	17	18	21	23	20
	V	18	18	16	19	20

Solution: The row-wise reductions are shown in Table 7.12

Table 7.12: Row-wise Reduction Matrix

		Men				
		1	2	3	4	5
Job	I	5	0	3	5	10
	II	6	8	0	2	3
	III	0	2	4	6	4
	IV	0	1	4	6	3
	V	2	2	0	3	4

The column wise reductions are shown in Table 7.13.

Table 7.13: Column-wise Reduction Matrix

		Men				
		1	2	3	4	5
Job	I	5	0	3	3	7
	II	6	8	0	0	0
	III	0	2	4	4	1
	IV	0	1	4	4	0
	V	2	2	0	1	1

Matrix with minimum number of lines drawn to cover all zeros is shown in Table 7.14.

Table 7.14: Matrix will all Zeros Covered

		Men				
		1	2	3	4	5
Job	I	5	0	3	3	7
	II	6	8	0	0	0
	III	0	2	4	4	1
	IV	0	1	4	4	0
	V	2	2	0	1	1

The number of lines drawn is 5, which is equal to the order of matrix. Hence optimality is reached. The optimal assignments are shown in Table 7.15.

Table 7.15: Optimal Assignment

	Men					
	1	2	3	4	5	
Job	I	5	0	3	3	7
	II	6	8	X	0	X
	III	0	2	4	4	1
	IV	X	1	4	4	0
	V	2	2	0	1	1

Therefore, the optimal solution is:

Job	Men	Time
I	2	15
II	4	14
III	1	21
IV	5	20
V	3	16
Total time = 86 hours		

7.7 UNBALANCED ASSIGNMENT PROBLEM

If the given matrix is not a square matrix, the assignment problem is called an **unbalanced problem**. In such type of problems, add dummy row(s) or column(s) with the cost elements as zero to convert the matrix as a square matrix. Then the assignment problem is solved by the Hungarian method.

Example 3: A company has five machines that are used for four jobs. Each job can be assigned to one and only one machine. The cost of each job on each machine is given in the following Table 7.16.

Table 7.16: Assignment Problem

	Machines					
	A	B	C	D	E	
Job	1	5	7	11	6	7
	2	8	5	5	6	5
	3	6	7	10	7	3
	4	10	4	8	2	4

Solution: Convert the 4×5 matrix into a square matrix by adding a dummy row D5.

Table 7.17: Dummy Row D5 Added

	Machines				
	A	B	C	D	E
1	5	7	11	6	7
2	8	5	5	6	5
Job 3	6	7	10	7	3
4	10	4	8	2	4
D ₅	0	0	0	0	0

Table 7.18: Row-wise Reduction of the Matrix

	Machines				
	A	B	C	D	E
1	0	2	6	1	2
2	3	0	0	1	0
Job 3	3	4	7	4	0
4	8	2	6	2	0
D ₅	0	0	0	0	0

Column-wise reduction is not necessary since all columns contain a single zero. Now, draw minimum number of lines to cover all the zeros, as shown in Table 7.19.

Table 7.19: All Zeros in the Matrix Covered

	Machines				
	A	B	C	D	E
1	0	2	6	1	2
2	3	0	0	1	0
Job 3	3	4	7	4	0
4	8	2	6	2	0
D ₅	0	0	0	0	0

Number of lines drawn ≠ Order of matrix. Hence not optimal.

Select the least uncovered element, i.e., 1, subtract it from other uncovered elements, add to the elements at intersection of lines and leave the elements that are covered with single line unchanged as shown in Table 7.20.

Table 7.20: Subtracted or Added to Elements

		Machines				
		A	B	C	D	E
Job	1	0	1	5	0	2
	2	4	0	0	1	1
	3	3	3	6	3	0
	4	8	1	5	1	0
	D ₅	1	0	0	0	1

Number of lines drawn ≠ Order of matrix. Hence not optimal.

Table 7.21: Again Added or Subtracted 1 from Elements

		Machines				
		A	B	C	D	E
Job	1	0	1	5	0	3
	2	4	0	0	1	2
	3	2	2	5	2	0
	4	7	0	4	0	0
	D ₅	1	0	0	0	2

Number of lines drawn = Order of matrix. Hence optimality is reached. Now assign the jobs to machines, as shown in Table 7.22.

Table 7.22: Assigning Jobs to Machines

		Machines				
		A	B	C	D	E
Job	1	0	1	5	X	3
	2	4	0	X	1	2
	3	2	2	5	2	0
	4	7	X	4	0	X
	D ₅	1	X	0	X	2

Hence, the optimal solution is:

Assignment Model

Job	Machine	Cost
1	A	5
2	B	5
3	E	3
4	D	2
D5	C	0
Total Cost		= Rs.15.00

Example 4: In a plant layout, four different machines M_1 , M_2 , M_3 and M_4 are to be erected in a machine shop. There are five vacant areas A, B, C, D and E. Because of limited space, Machine M_2 cannot be erected at area C and Machine M_4 cannot be erected at area A. The cost of erection of machines is given in the Table 7.23.

Table 7.23: Assignment Problem

		Area				
		A	B	C	D	E
Machine	M_1	4	5	9	4	5
	M_2	6	4	--	4	3
	M_3	4	5	8	5	1
	M_4	--	2	6	1	2

Find the optimal assignment plan.

Solution: As the given matrix is not balanced, add a dummy row D_5 with zero cost values. Assign a high cost H for (M_2, C) and (M_4, A) . While selecting the lowest cost element neglect the high cost assigned H, as shown in Table 7.24 below.

Table 7.24: Dummy Row D_5 Added

		Area				
		A	B	C	D	E
Machine	M_1	4	5	9	4	5
	M_2	6	4	H	4	3
	M_3	4	5	8	5	1
	M_4	H	2	6	1	2
		D_5	0	0	0	0

Row-wise reduction of the matrix, is shown in Table 7.25.

Table 7.25: Matrix Reduced Row-wise Area

	A	B	C	D	E
M₁	0	1	5	0	1
Machine	M₂	3	1	H	1
	M₃	3	4	7	4
	M₄	H	1	5	0
	D₅	0	0	0	0

Note: Column-wise reduction is not necessary, as each column has at least one single zero. Now, draw minimum number of lines to cover all the zeros, see Table 7.26.

Table 7.26: Lines Drawn to Cover all Zeros Area

	A	B	C	D	E
M₁	0	1	5	0	1
Machine	M₂	3	1	H	1
	M₃	3	4	7	4
	M₄	H	1	5	0
	D₅	0	0	0	0

Number of lines drawn \neq Order of matrix. Hence not Optimal. Select the smallest uncovered element, in this case 1. Subtract 1 from all other uncovered element and add 1 with the elements at the intersection. The element covered by single line remains unchanged. These changes are shown in Table 7.27. Now try to draw minimum number of lines to cover all the zeros.

Table 7.27: Added or Subtracted 1 from Elements

	A	B	C	D	E
M₁	0	1	5	1	2
Machine	M₂	2	0	H	1
	M₃	2	3	6	4
	M₄	H	0	4	0
	D₅	0	0	0	1

Now number of lines drawn = Order of matrix, hence optimality is reached. Optimal assignment of machines to areas are shown in Table 7.28.

Table 7.28: Optimal Assignment

		Area				
		A	B	C	D	E
Machine	M ₁	0	1	5	1	2
	M ₂	2	0	H	1	X
	M ₃	2	3	6	4	0
	M ₄	H	0	4	0	1
	D ₅	X	X	0	1	1

Hence, the optimal solution is:

Machines	Area	Erection Cost
M ₁	A	4
M ₂	B	4
M ₃	C	1
M ₄	D	1
D ₅	E	0
Total Erection Cost = Rs.10.00		

7.8 RESTRICTED ASSIGNMENT PROBLEM

In real practice, situations may arise where a particular machine cannot be assigned to an operator because he may not be skilled enough to operate it. Because of this, no assignment is made for the operator on that machine. This situation is overcome by assigning a large value, or by assigning M. This will result in no assignment made to the restricted combinations.

Example 5: Five jobs are to be assigned to five men. The cost (in Rs.) of performing the jobs by each man is given in the matrix (Table 7.29). The assignment has restrictions that Job 4 cannot be performed by Man 1 and Job 3 cannot be performed by Man 4. Find the optimal assignment of job and its cost involved.

Table 7.29: Assignment Problem

Men	1	2	3	4	5
1	16	12	11	x	15
2	13	15	11	16	18
3	20	21	18	19	17
4	16	13	x	16	12
5	20	19	18	17	19

Solution: Assign large value to the restricted combinations or introduce 'M', see Table 7.30.

Table 7.30: Large Value Assignment to Restricted Combinations

		Job				
		1	2	3	4	5
Men	1	16	12	11	M	15
	2	13	15	11	16	18
	3	20	21	18	19	17
	4	16	13	M	16	12
	5	20	19	18	17	19

Table 7.31: Reducing the matrix row-wise

		Job				
		1	2	3	4	5
Men	1	5	1	0	M	4
	2	2	4	0	5	7
	3	3	4	1	2	0
	4	4	1	M	4	0
	5	3	2	1	0	1

Table 7.32: Reducing the matrix column-wise

		Job				
		1	2	3	4	5
Men	1	3	0	0	M	4
	2	0	3	0	5	7
	3	1	3	1	2	0
	4	2	0	M	4	0
	5	1	1	1	0	1

Draw minimum number of lines to cover all zeros, see Table 7.33.

Table 7.33: All Zeros Covered

		Job				
		1	2	3	4	5
Men	1	-3	0	0	M	+
	2	0	-3	0	5	7
	3	1	3	1	2	0
	4	-2	0	M	4	0
	5	1	1	1	0	1

Now, number of lines drawn = Order of matrix, hence optimality is reached (Table 7.34).

Assignment Model

Allocating Jobs to Men.

Table 7.34: Job Allocation to Men

		Job				
		1	2	3	4	5
Men	1	3	X	0	M	4
	2	0	3	X	5	7
	3	1	3	1	2	0
	4	2	0	M	4	X
	5	1	1	1	0	1

Table 7.35: Assignment Schedule and Cost

Men	Job	Cost
1	3	11
2	1	13
3	5	17
4	2	13
5	4	17
		Total Cost = Rs. 71.00

As per the restriction conditions given in the problem, Man 1 and Man 4 are not assigned to Job 4 and Job 3 respectively.

7.9 MULTIPLE AND UNIQUE SOLUTIONS

For a given Job-Men assignment problem, there can be more than one optimal solution, i.e., multiple solutions can exist. Two assignment schedules that give same results are called ***Multiple optimal solutions***. If the problem has only one solution then the solution is said to be ***Unique solution***. A problem having multiple optimal solutions is shown in Example 4.6.

7.10 MAXIMIZATION PROBLEM

In maximization problem, the objective is to maximize profit, revenue, etc. Such problems can be solved by converting the given maximization problem into a minimization problem.

- Change the signs of all values given in the table.
- Select the highest element in the entire assignment table and subtract all the elements of the table from the highest element.

Example 6: A marketing manager has five salesmen and sales districts. Considering the capabilities of the salesmen and the nature of districts, the marketing manager estimates that sales per month (in hundred rupees) for each salesman in each district would be as follows (Table 7.36). Find the assignment of salesmen to districts that will result in maximum sales.

Table 7.36: Maximization Problem

		District				
		A	B	C	D	E
Salesman	1	32	38	40	28	40
	2	40	24	28	21	36
	3	41	27	33	30	37
	4	22	38	41	36	36
	5	29	33	40	35	39

Solution: The given maximization problem is converted into minimization problem (Table 7.37) by subtracting from the highest sales value (i.e., 41) with all elements of the given table.

Table 7.37: Conversion to Minimization Problem

		District				
		A	B	C	D	E
Salesman	1	9	3	1	13	1
	2	1	17	13	20	5
	3	0	14	8	11	4
	4	19	3	0	5	5
	5	12	8	1	6	2

Reduce the matrix row-wise (see Table 7.38)

Table 7.38: Matrix Reduced Row-wise

		District				
		A	B	C	D	E
Salesman	1	8	2	0	12	0
	2	0	16	12	19	4
	3	0	14	8	11	4
	4	19	3	0	5	5
	5	11	7	0	5	1

Reduce the matrix column-wise and draw minimum number of lines to cover all the zeros in the matrix, as shown in Table 7.39.

Table 7.39: Matrix Reduced Column-wise and Zeros Covered

		District				
		A	B	C	D	E
Salesman	1	8	0	0	7	0
	2	0	14	12	14	4
	3	0	12	8	6	4
	4	19	1	0	0	5
	5	11	5	0	0	1

Number of lines drawn \neq Order of matrix. Hence not optimal.

Select the least uncovered element, i.e., 4 and subtract it from other uncovered elements, add it to the elements at intersection of line and leave the elements that are covered with single line unchanged, Table 7.40.

Table 7.40: Added & Subtracted the least Uncovered Element

		District				
		A	B	C	D	E
Salesman	1	12	0	0	7	0
	2	0	10	8	10	0
	3	0	8	4	2	0
	4	23	1	0	0	5
	5	15	5	0	0	1

Now, number of lines drawn = Order of matrix, hence optimality is reached.

There are two alternative assignments due to presence of zero elements in cells (4, C), (4, D), (5, C) and (5, D).

Table 7.41: Two Alternative Assignments

		A	B	C	D	E			A	B	C	D	E
Salesman	1	12	0	X	7	X	1	12	0	X	7	X	
	2	X	10	8	10	X	2	X	10	8	10	0	
	3	0	8	4	2	0	3	0	8	4	2	X	
	4	23	1	0	X	5	4	23	1	0	X	5	
	5	15	5	X	0	1	5	15	5	X	0	1	

Therefore,

Assignment 1			Assignment 2		
Salesman	Districts	Sales (in '00) Rs.	Salesman	Districts	Sales (in '00) Rs.
1	B	38	1	B	38
2	A	40	2	E	36
3	E	37	3	A	41
4	C	41	4	C	41
5	D	35	5	D	35
Total Rs. = 191.00			Total Rs. = 191.00		

7.11 TRAVELLING SALESMAN PROBLEM

The ‘Travelling salesman problem’ is very similar to the assignment problem except that in the former, there are additional restrictions that a salesman starts from his city, visits each city once and returns to his home city, so that the total distance (cost or time) is minimum.

Procedure:

- Step 1: Solve the problem as an assignment problem.
- Step 2: Check for a complete cycle or alternative cycles. If the cycle is complete, Go to Step 4. If not, go to the Step 3.
- Step 3: To start with, assign the next least element other than zero, (only for first allocation) and complete the assignment. Go to Step 2.
- Step 4: Write the optimum assignment schedule and calculate the cost/time.

(**Note:** If there are two non-zero values in the matrix, it means that there are two optimal solutions. Calculate the cost for the two allocations and find the optimal solution.)

Example 7: A Travelling salesman has to visit five cities. He wishes to start from a particular city, visit each city once and then return to his starting point. The travelling cost (in Rs.) of each city from a particular city is given below.

Table 7.42: Travelling Salesman Problem

		To city				
		A	B	C	D	E
From city	A	α	2	5	7	1
	B	6	α	3	8	2
C	8	7	α	4	7	
D	12	4	6	α	5	
E	1	3	2	8	α	

What should be the sequence of the salesman's visit, so that the cost is minimum?

Solution: The problem is solved as an assignment problem using Hungarian method; an optimal solution is reached as shown in Table 7.43.

Table 7.43: Optimal Solution Reached Using Hungarian Method

		To city				
		A	B	C	D	E
From city	A	α	1	3	6	0
	B	4	α	0	6	X
	C	4	3	α	0	3
	D	8	0	1	α	1
	E	0	2	X	7	α

In this assignment, it means that the travelling salesman will start from city A, then go to city E and return to city A without visiting the other cities. The cycle is not complete. To overcome this situation, the next highest element can be assigned to start with. In this case it is 1, and there are three 1's. Therefore, consider all these 1's one by one and find the route which completes the cycle.

Case 1: Make the assignment for the cell (A, B) which has the value 1. Now, make the assignments for zeros in the usual manner. The resulting assignments are shown in Table 7.44.

Table 7.44: Resulting Assignment

		To city				
		A	B	C	D	E
From city	A	α	1	3	6	X
	B	4	α	0	6	X
	C	4	3	α	0	3
	D	8	X	1	α	1
	E	0	2	X	7	α

The assignment shown in Table 7.42 gives the route sequence

$A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow E$ and $E \rightarrow A$.

The travelling cost to this solution is

$$= 2000 + 3000 + 4000 + 5000 + 1000$$

$$= \text{Rs.}15,000.00$$

Case 2: If the assignment is made for cell (D, C) instead of (D, E), the feasible solution cannot be obtained. The route for the assignment will be $A \rightarrow B \rightarrow C \rightarrow D \rightarrow C$. In this case, the salesman visits city C twice and cycle is not complete.

Therefore the sequence feasible for this assignment is

$A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow A$.

with the travelling cost of Rs.15,000.00

7.12 SOLVING PROBLEMS ON THE COMPUTER WITH TORA

Transportation model option is used for assignment values. Similar to transportation model the cost or time values are entered in the input grid. With the constraint that each operator has to be assigned with one job, the supply and demand values are entered as 1. For example, the worked out example 1 is used for solving using computer.

Input screen:

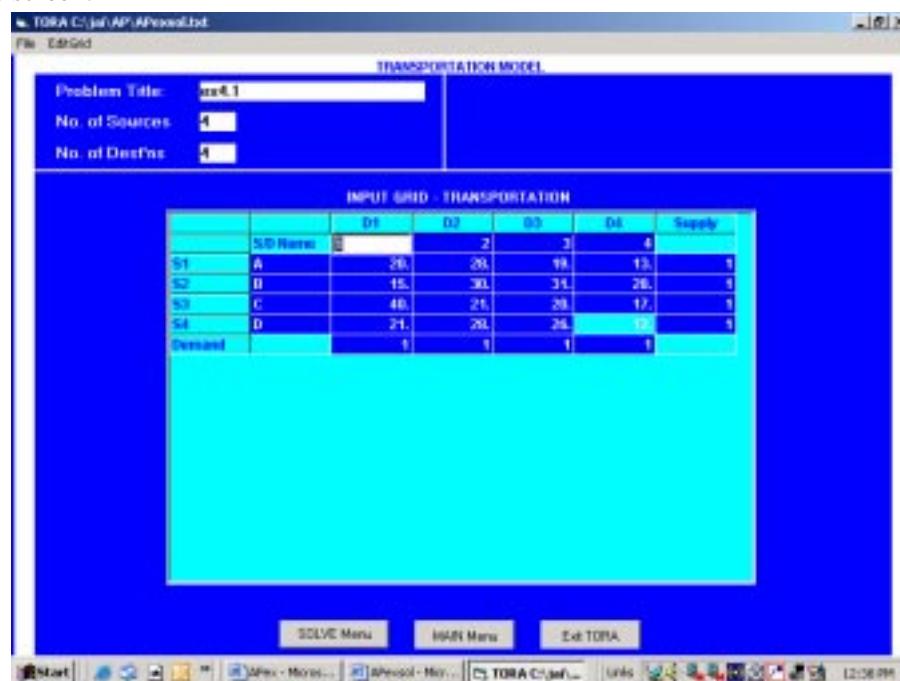


Figure 7.4: Assignment Problem Using TORA (Input Screen)

Output screen:

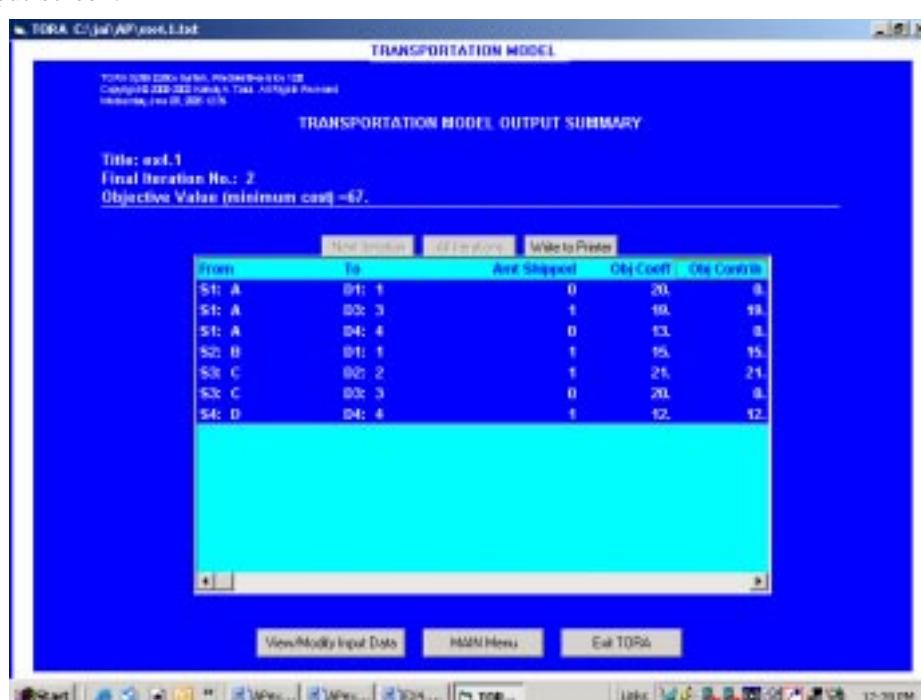


Figure 7.5: Assignment Problem Using TORA (Output Screen)

From the output screen, the objective is to minimize cost = Rs. 67.00

The assignment schedule is given below in Table 7.45.

Table 7.45: Assignment Schedule

Task	Operator	Cost
A	3	19
B	1	15
C	2	21
D	4	12
Total Cost = Rs.67.00		

7.13 SOLVING UNBALANCED ASSIGNMENT PROBLEM USING COMPUTER

Worked out Example 3 has been solved again using computer. The Input screen 4×5 matrix is shown.

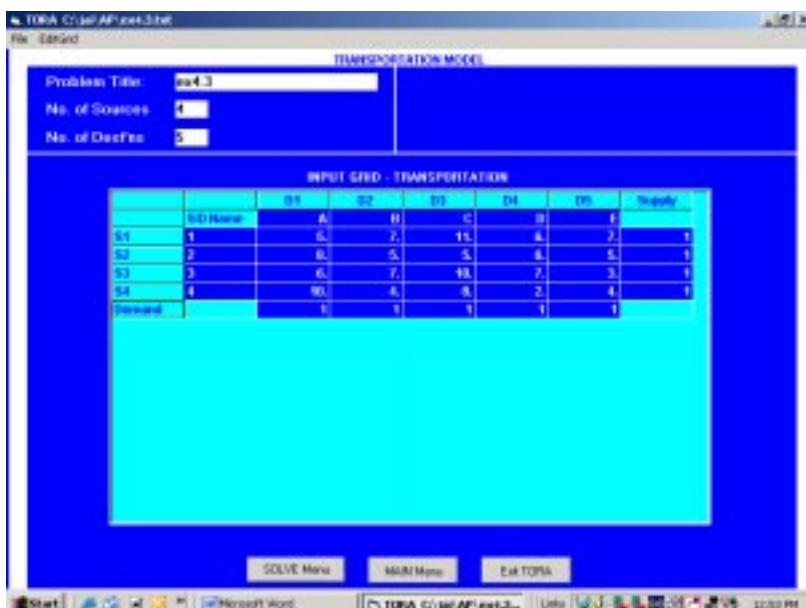


Figure 7.6: Unbalanced Assignment Problems Using TORA (Input Screen)

Output screen:

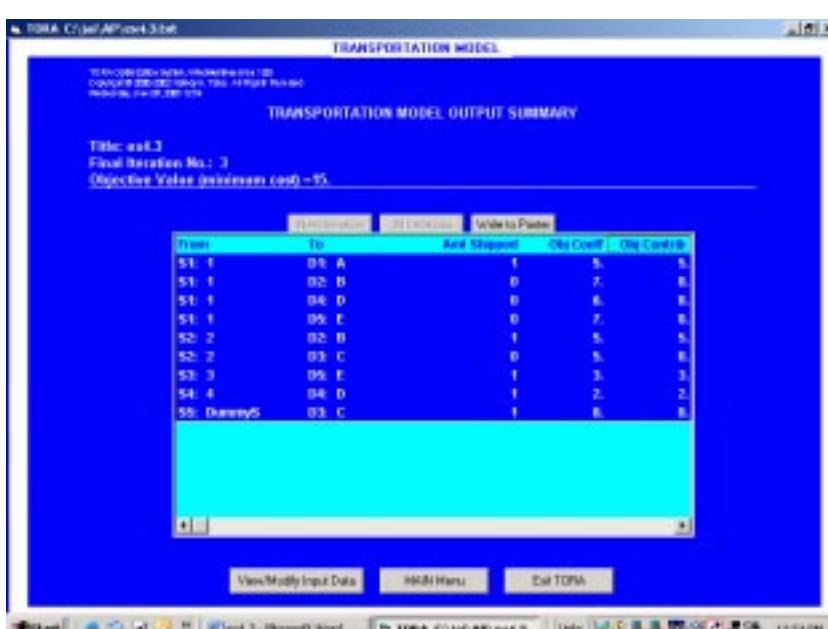


Figure 7.7: Unbalanced Assignment Problem Using TORA (Output Screen)

From the output obtained, the objective function value is Rs.15.00.

The assignment schedule is given in the Table 7.46 below.

Table 7.46: Assignment Schedule

Job	Machine	Cost
1	A	5
2	B	5
3	E	3
4	D	2
D5	C	0
		Total Cost = Rs. 15.00

7.14 SOLVING MAXIMIZATION PROBLEMS USING COMPUTERS

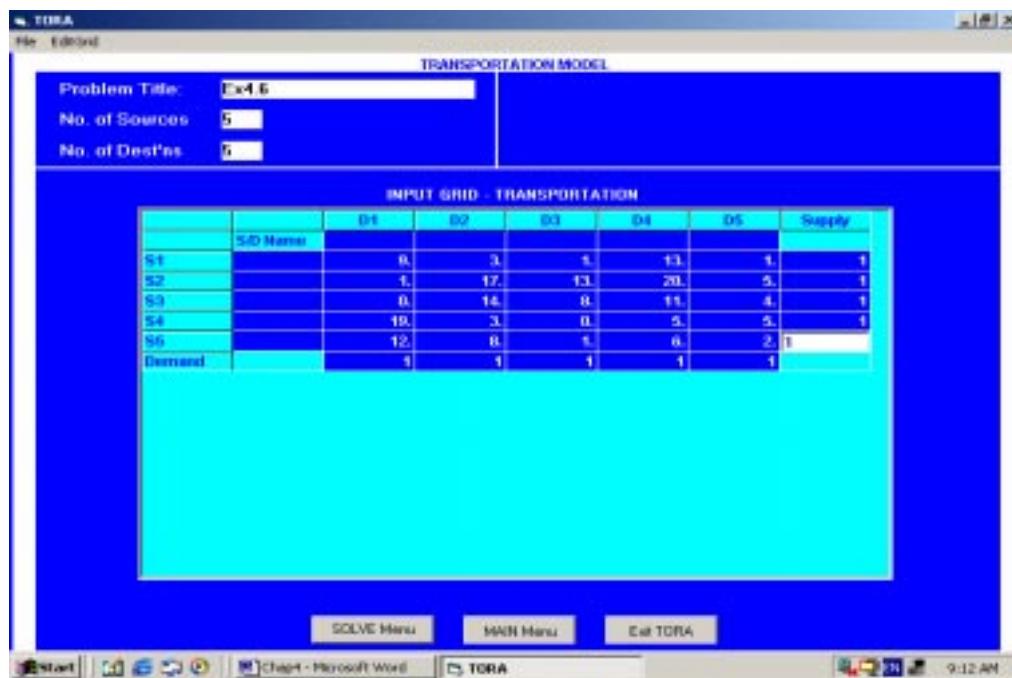
As we know, the transportation model is also used for solving assignment problems. In transportation model, the objective is to minimize the cost of transportation. For a maximization problem, the objective is to maximize the profit or returns. While entering the values the maximization matrix must be converted to minimization matrix by subtracting all the values with the highest value cell. This is shown by solving the solved problem Ex. 6. The given problem is maximization of sales (Table 7.47).

Table 7.47: Maximization Problem

		District					
		A	B	C	D	E	
Salesman		1	32	38	40	28	40
		2	40	24	28	21	36
Salesman	3	41	27	33	30	37	
Salesman	4	22	38	41	36	36	
Salesman	5	29	33	40	35	39	

Taking the highest value in the given maximization matrix, i.e., 41 and subtracting all other values, we get the following input matrix:

		District					
		A	B	C	D	E	
Salesman		1	9	3	1	13	1
		2	1	17	13	20	5
Salesman	3	0	14	8	11	4	
Salesman	4	19	3	0	5	5	
Salesman	5	12	8	1	6	2	

**Figure 7.8: Solving Maximization Using TORA (Input Screen)**

Part of the output screen is shown below in Figure 7.9.

From	To	Unit Shipped
S1:	D2:	1
S2:	D1:	0
S2:	D5:	1
S3:	D1:	1
S3:	D4:	0
S4:	D2:	0
S4:	D3:	1
S5:	D3:	0
S5:	D4:	1

Figure 7.9: Part of Output Screen (Enlarged)

The output given by TORA is the assignment schedule with the objective of minimization. The given problem is to maximize the sales. To arrive at the maximize sales value, add the assigned values from the given matrix, as shown in Table 7.48.

Table 7.48: Assignment Schedule

Salesman	District	*Sales (in '00) Rs.
1	B	38
2	E	36
3	A	41
4	C	41
5	D	35
		Total Cost = Rs.191.00

* values taken from the given matrix.

Check Your Progress 7.1

1. How could an assignment problem be solved using the transportation approach.
2. Describe the approach you would use to solve an assignment problem with the help of illustration.

Notes: (a) Write your answer in the space given below.
(b) Please go through the lesson sub-head thoroughly you will get your answers in it.
(c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

7.15 LET US SUM UP

AP brings into play the allocation of a number of jobs to a number of persons in order to minimize the completion time. Although an AP can be formulated as LPP, it is solved by a special method known as Hungarian method. The Hungarian method of assignment provides us with an efficient means of finding the optimal solution without having to make a direct comparison of every option. Further we will take into consideration the opportunity cost. This is a next best alternative cost.

7.16 LESSON-END ACTIVITY

Visit to your nearest fast moving consumer goods manufacturing company like LG, Samsung, Videocon, Onida etc. and apply the concept of assignment model to increase its production line.

7.17 KEYWORDS

Balanced Assignment Problem

Unbalanced Assignment Problem

Hungarian Method

Restricted Assignment Problem

Dummy job

Opportunity cost

7.18 QUESTIONS FOR DISCUSSION

1. Write True or False against each statement:

- (a) Basic objective of an AP is to assign n-number of resources to a number of activities.
- (b) Application of AP is an allocation of machine for optimum utilization of space.

- (c) Hungarian method could also be applicable to transportation model.
- (d) Assignment problem not consider the allocation of number of jobs to a number of person.
- (e) An optimal assignment is found, if the number of assigned cells equal the number of row (columns).
- 2. Briefly comment on the following statement:**
- Assignment problem are of the types balanced and unbalanced.
 - Cost or time value for the dummy cells are assumed zero.
 - Maximization problem objective is to maximize profit.
- 3. Fill in the blank:**
- Assignment model can be applied in many _____.
 - If the given matrix is not a _____, matrix, the AP is called an _____ problem.
 - Transportation model is used for _____ values.
 - A dummy job is an _____ jobs.
- 4. Write short Notes:**
- What is meant by matrix reduction.
 - Describe the approach of the Hungarian method.

7.19 TERMINAL QUESTIONS

- What is an assignment problem? Give its areas of application.
- Explain the structure of an assignment problem with objectives as maximization and minimization.
- How can an assignment problem be solved using linear programming ? Illustrate with a suitable example.
- Explain the steps involved in solving an assignment problem.
- What is meant by an unbalanced assignment problem?
- How is an assignment problem solved when certain assignments are restricted?
- What is the difference between a multiple and a unique solution in an assignment problem?
- How is a maximization problem dealt with, in solving assignment problems?
- What is Travelling-salesman problem? How does it differ from an assignment problem?
- Discuss how assignment problems are solved using transportation model.

Exercise Problems

- Consider the assignment problem having the following cost table:

Men	Job		
	1	2	3
A	7	9	6
B	5	8	7
C	4	5	6

- Draw the network representation of the problem.
- Solve the problem and determine the optimal assignment for each man.

2. Consider the assignment problem having the following table. Use TORA to find the optimal solution that minimizes the total cost:

Operator	Job				
	1	2	3	4	5
A	12	14	16	11	10
B	9	13	17	9	7
C	10	12	20	7	8
D	13	10	21	6	12
E	15	9	15	11	13

3. Four trucks are used for transporting goods to four locations. Because of varying costs of loading and unloading the goods, the cost of transportation also varies for each truck. The cost details (in Rs.) is given in the table below. There is no constraint, and any truck can be sent to any location. The objective is to assign the four trucks to minimize the total transportation cost. Formulate and solve the problem using TORA.

Truck	Location			
	1	2	3	4
A	525	825	320	200
B	600	750	250	175
C	500	900	270	150
D	620	800	300	160

4. A two-wheeler service station head has four workmen and four tasks to be performed daily as a routine work. Before assigning the work, the service station head carried out a test by giving each work to all the workmen. The time taken by workmen is given in the table, below.

Work	Time Taken (in mins)			
	Workman			
	1	2	3	4
A	20	28	19	13
B	15	30	16	23
C	40	17	20	13
D	17	28	22	8

How should the service station head assign the work to each workman so as to minimize the total time?

5. Consider an unbalanced assignment problem having the following cost table:

Operator	Task			
	1	2	3	4
A	12	14	15	16
B	10	11	13	21
C	8	9	17	23

6. Consider the following assignment problem:

Assignment Model

Destination	Unit cost (Rs.)				Supply
	1	2	3	4	
Source					
1	30	61	45	50	1
2	25	54	49	52	1
3	27	60	45	54	1
4	31	57	49	55	1
Demand	1	1	1	1	

- a. Draw the network representation of the assignment problem.
 - b. Formulate a linear programming model for the assignment problem.
7. Five operators have to be assigned to five machines. Depending on the efficiency and skill, the time taken by the operators differs. Operator B cannot operate machine 4 and operator D cannot operate machine 2. The time taken is given in the following table.

Operator	Machine				
	1	2	3	4	5
A	6	6	3	---	5
B	6	7	2	5	3
C	5	6	4	6	4
D	7	---	7	6	7
E	5	4	3	6	5

Determine the optimal assignment using TORA.

8. A consumer durables manufacturing company has plans to increase its product line, namely, washing machine, refrigerator, television and music system. The company is setting up new plants and considering four locations. The demand forecast per month for washing machine, refrigerator, television and music system are 1000, 750, 850 and 1200, respectively. The company decides to produce the forecasted demand. The fixed and variable cost per unit for each location and item is given in the following table. The management has decided not to set-up more than one unit in one location.

Location	Fixed cost (lakhs)				Variable cost / unit			
	WM	RF	TV	MS	WM	RF	TV	MS
Chennai	30	35	18	16	4	3	6	2
Coimbatore	25	40	16	12	3	2	4	4
Madurai	35	32	15	10	4	2	7	6
Selam	20	25	14	12	2	1	3	7

Determine the location and product combinations so that the total cost is minimized.

9. Solve the following travelling salesman problem so as to minimize the cost of travel.

		City				
		A	B	C	D	E
A	A	--	13	22	21	11
	B	2	--	11	16	3
C	9	9	--	20	10	
D	13	12	27	--	16	
E	12	10	28	26	--	

10. Solve the travelling salesman problem for the given matrix cell values which represent the distances between cities.

$$c_{12} = 31, \quad c_{13} = 10, \quad c_{14} = 15,$$

$$c_{21} = 9, \quad c_{23} = 12, \quad c_{31} = 10,$$

$$c_{34} = 9, \quad c_{41} = 18, \quad c_{42} = 25.$$

There is no route between cities i and j if value for c_{ij} is not given.

7.20 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

1. (a) True (b) True (c) False (d) False (e) True
 3. (a) decision-making (b) square, unbalanced
 (c) assignment (d) imaginary

7.21 SUGGESTED READINGS

Ross, G.T. and Soaland, R.H, "Modeling facility location problem as generalized assignment problems", Management Science.

U.L. Gupta, D.T. Lee, J.T. Leung, An optimal solution for the channel-assignment problem.

Abara J., Applying Integer Linear Programming to the Fleet Assignment Problem, Interfaces, Vol. 19, No. 41, pp. 20-28.

Unit-III

LESSON

8

NETWORK MODEL

CONTENTS

- 8.0 Aims and Objectives
- 8.1 Introduction
- 8.2 PERT / CPM Network Components
- 8.3 Errors to be avoided in Constructing a Network
- 8.4 Rules in Constructing a Network
- 8.5 Procedure for Numbering the Events Using Fulkerson's Rule
- 8.6 Critical Path Analysis
- 8.7 Determination of Float and Slack Times
- 8.8 Solving CPM Problems using Computer
- 8.9 Project Evaluation Review Technique, PERT
- 8.10 Solving PERT Problems using Computer
- 8.11 Cost Analysis
- 8.12 Let us Sum Up
- 8.13 Lesson-end Activity
- 8.14 Keywords
- 8.15 Questions for Discussion
- 8.16 Terminal Questions
- 8.17 Model Answers to Questions for Discussion
- 8.18 Suggested Readings

8.0 AIMS AND OBJECTIVES

In this lesson we are going to discuss the various Network Model Like Critical Path Method and Project Evaluations Review Technique. The CPM in a diagrammatical technique whereas PERT in a unique controlling device.

8.1 INTRODUCTION

Any project involves planning, scheduling and controlling a number of interrelated activities with use of limited resources, namely, men, machines, materials, money and time. The projects may be extremely large and complex such as construction of a power plant, a highway, a shopping complex, ships and aircraft, introduction of new products and research and development projects. It is required that managers must have a dynamic planning and scheduling system to produce the best possible results and also to react immediately to the changing conditions and make necessary changes in the plan and schedule. A convenient analytical and visual technique of PERT and CPM prove extremely valuable in assisting the managers in managing the projects.

Both the techniques use similar terminology and have the same purpose. PERT stands for **Project Evaluation and Review Technique** developed during 1950's. The technique was developed and used in conjunction with the planning and designing of the Polaris missile project. CPM stands for **Critical Path Method** which was developed by DuPont Company and applied first to the construction projects in the chemical industry. Though both PERT and CPM techniques have similarity in terms of concepts, the basic difference is, PERT is used for analysis of project scheduling problems. CPM has single time estimate and PERT has three time estimates for activities and uses probability theory to find the chance of reaching the scheduled time.

Project management generally consists of three phases.

Planning: Planning involves setting the objectives of the project. Identifying various activities to be performed and determining the requirement of resources such as men, materials, machines, etc. The cost and time for all the activities are estimated, and a network diagram is developed showing sequential interrelationships (predecessor and successor) between various activities during the planning stage.

Scheduling: Based on the time estimates, the start and finish times for each activity are worked out by applying forward and backward pass techniques, critical path is identified, along with the slack and float for the non-critical paths.

Controlling: Controlling refers to analyzing and evaluating the actual progress against the plan. Reallocation of resources, crashing and review of projects with periodical reports are carried out.

8.2 PERT/CPM NETWORK COMPONENTS

PERT / CPM networks contain two major components

- i. Activities, and
- ii. Events

Activity: An activity represents an action and consumption of resources (time, money, energy) required to complete a portion of a project. Activity is represented by an arrow, (Figure 8.1).

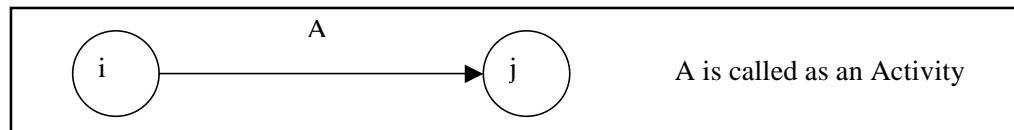


Figure 8.1: An Activity

Event: An event (or node) will always occur at the beginning and end of an activity. The event has no resources and is represented by a circle. The i^{th} event and j^{th} event are the tail event and head event respectively, (Figure 8.2).

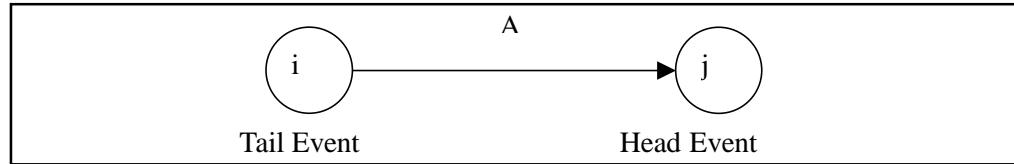


Figure 8.2: An Event

Merge and Burst Events

One or more activities can start and end simultaneously at an event (Figure 8.3 a, b).



Figure 8.3

Preceding and Succeeding Activities

Network Model

Activities performed before given events are known as *preceding activities* (Figure 8.4), and activities performed after a given event are known as *succeeding activities*.

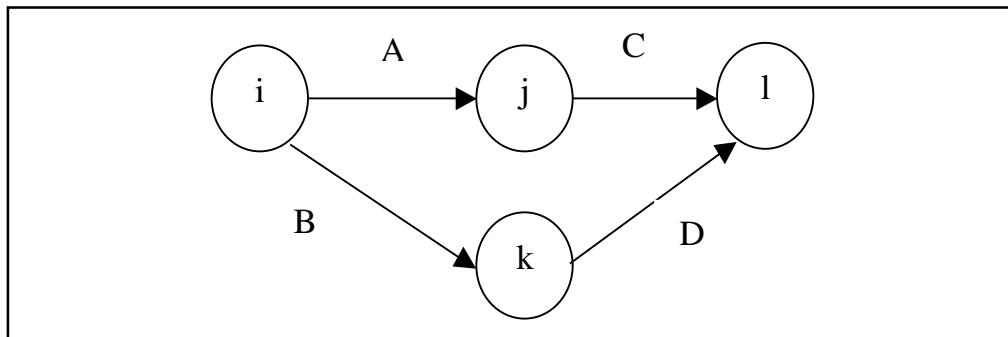


Figure 8.4: Preceding and Succeeding Activities

Activities A and B precede activities C and D respectively.

Dummy Activity

An imaginary activity which does not consume any resource and time is called a *dummy activity*. Dummy activities are simply used to represent a connection between events in order to maintain a logic in the network. It is represented by a dotted line in a network, see Figure 8.5.

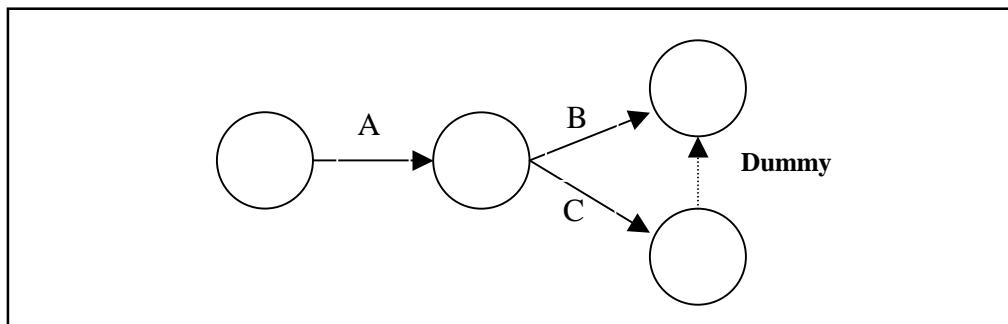


Figure 8.5: Dummy Activity

8.3 ERRORS TO BE AVOIDED IN CONSTRUCTING A NETWORK

- Two activities starting from a tail event must not have a same end event. To ensure this, it is absolutely necessary to introduce a dummy activity, as shown in Figure 8.6.

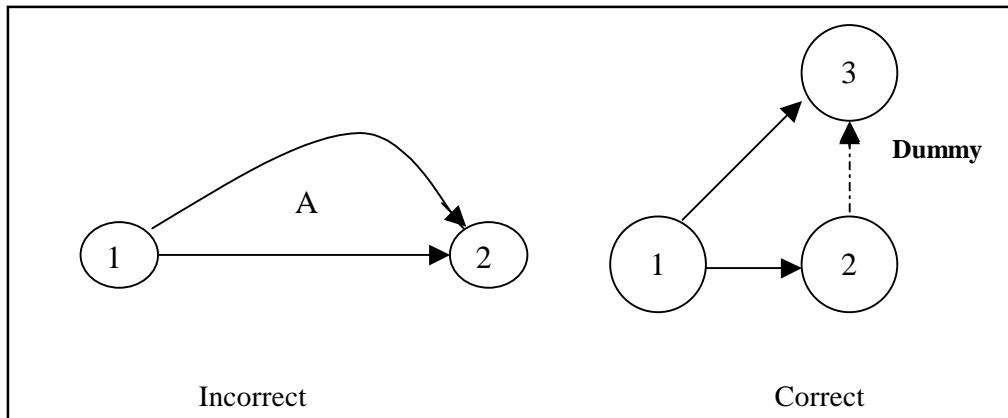


Figure 8.6: Correct and Incorrect Activities

- b. **Looping error** should not be formed in a network, as it represents performance of activities repeatedly in a cyclic manner, as shown below in Figure 8.7.

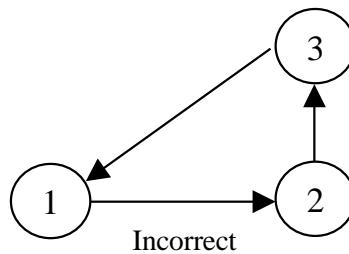


Figure 8.7: Looping Error

- c. In a network, there should be only one start event and one ending event as shown below, in Figure 8.8.

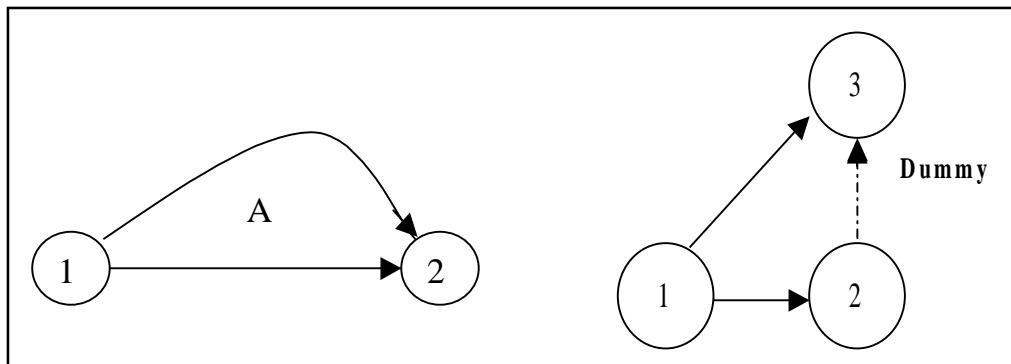


Figure 8.8: Only One Start and End Event

- d. The direction of arrows should flow from left to right avoiding mixing of direction as shown in Figure 8.9.

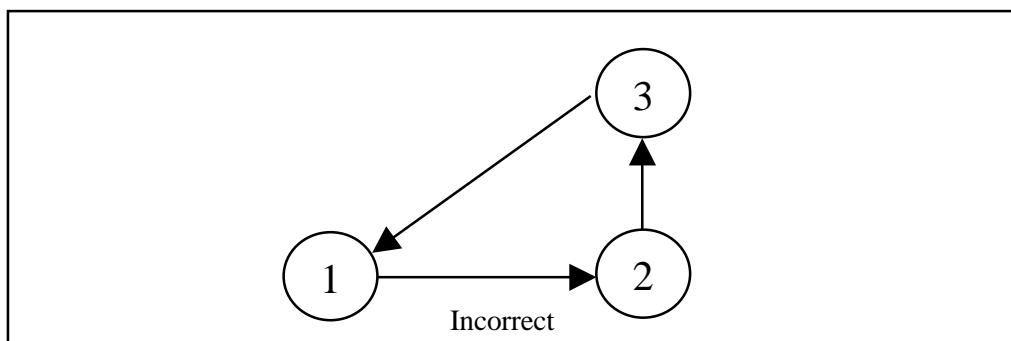


Figure 8.9: Wrong Direction of Arrows

8.4 RULES IN CONSTRUCTING A NETWORK

1. No single activity can be represented more than once in a network. The length of an arrow has no significance.
2. The event numbered 1 is the start event and an event with highest number is the end event. Before an activity can be undertaken, all activities preceding it must be completed. That is, the activities must follow a logical sequence (or – inter-relationship) between activities.
3. In assigning numbers to events, there should not be any duplication of event numbers in a network.
4. Dummy activities must be used only if it is necessary to reduce the complexity of a network.
5. A network should have only one start event and one end event.

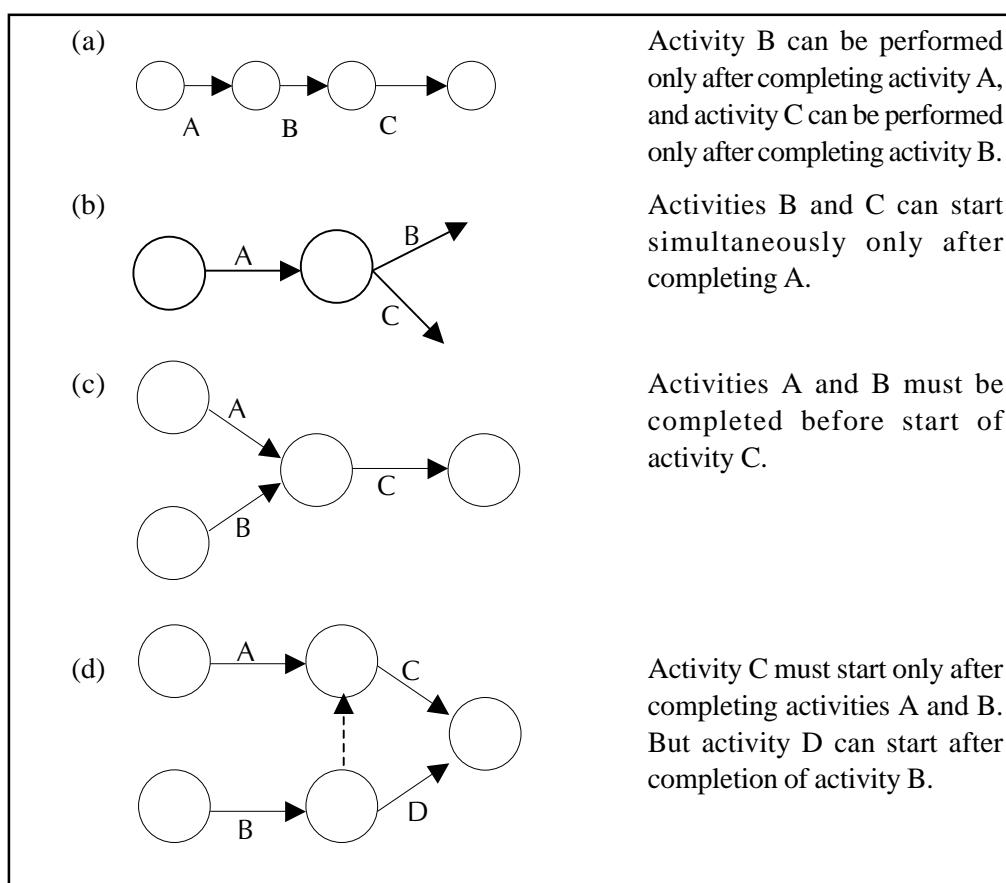


Figure 8.10 (a), (b), (c), (d): Some Conventions followed in making Network Diagrams

8.5 PROCEDURE FOR NUMBERING THE EVENTS USING FULKERSON'S RULE

- Step1:** Number the start or initial event as 1.
- Step2:** From event 1, strike off all outgoing activities. This would have made one or more events as initial events (event which do not have incoming activities). Number that event as 2.
- Step3:** Repeat step 2 for event 2, event 3 and till the end event. The end event must have the highest number.

Example 1: Draw a network for a house construction project. The sequence of activities with their predecessors are given in Table 8.1, below.

Table 8.1: Sequence of Activities for House Construction Project

Name of the activity	Starting and finishing event	Description of activity	Predecessor	Time duration (days)
A	(1,2)	Prepare the house plan	--	4
B	(2,3)	Construct the house	A	58
C	(3,4)	Fix the door / windows	B	2
D	(3,5)	Wiring the house	B	2
E	(4,6)	Paint the house	C	1
F	(5,6)	Polish the doors / windows	D	1

Solution:

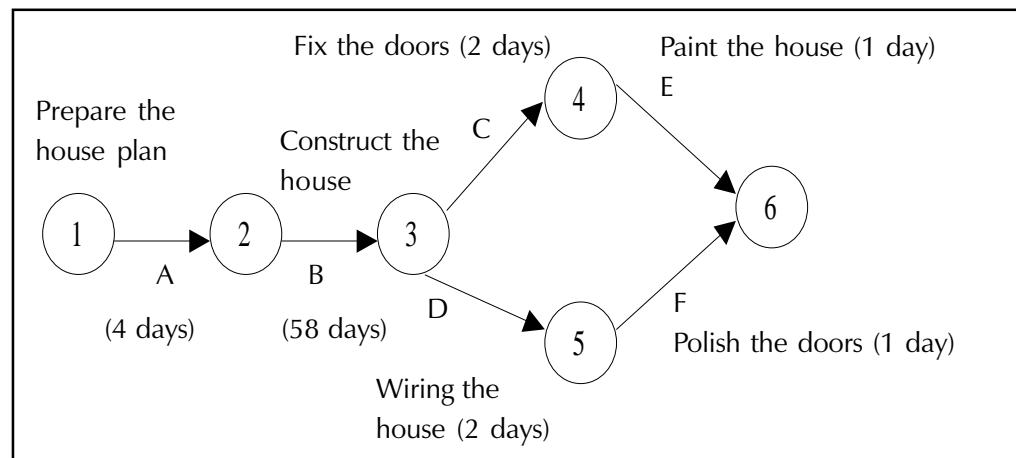


Figure 8.11: Network diagram representing house construction project.

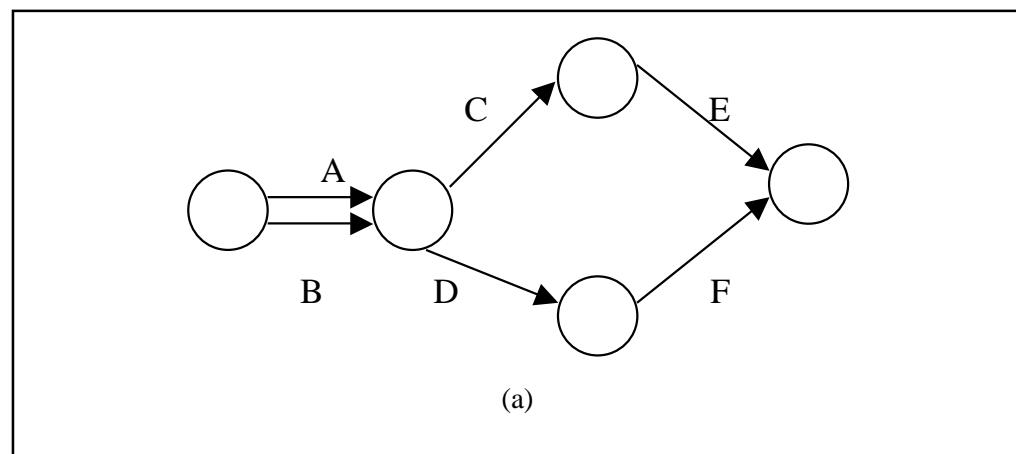
The network diagram in Figure 8.11 shows the procedure relationship between the activities. Activity A (preparation of house plan), has a start event 1 as well as an ending event 2. Activity B (Construction of house) begins at event 2 and ends at event 3. The activity B cannot start until activity A has been completed. Activities C and D cannot begin until activity B has been completed, but they can be performed simultaneously. Similarly, activities E and F can start only after completion of activities C and D respectively. Both activities E and F finish at the end of event 6.

Example 2: Consider the project given in Table 8.2 and construct a network diagram.

Table 8.2: Sequence of Activities for Building Construction Project

Activity	Description	Predecessor
A	Purchase of Land	-
B	Preparation of building plan	-
C	Level or clean the land	A
D	Register and get approval	A, B
E	Construct the building	C
F	Paint the building	D

Solution: The activities C and D have a common predecessor A. The network representation shown in Figure 8.12 (a), (b) violates the rule that no two activities can begin and end at the same events. It appears as if activity B is a predecessor of activity C, which is not the case. To construct the network in a logical order, it is necessary to introduce a dummy activity as shown in Figure 8.12.



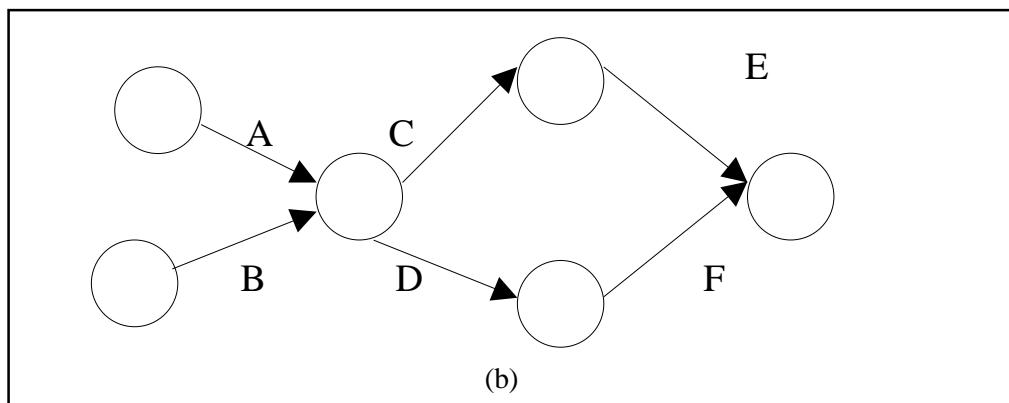


Figure 8.12: Network representing the Error

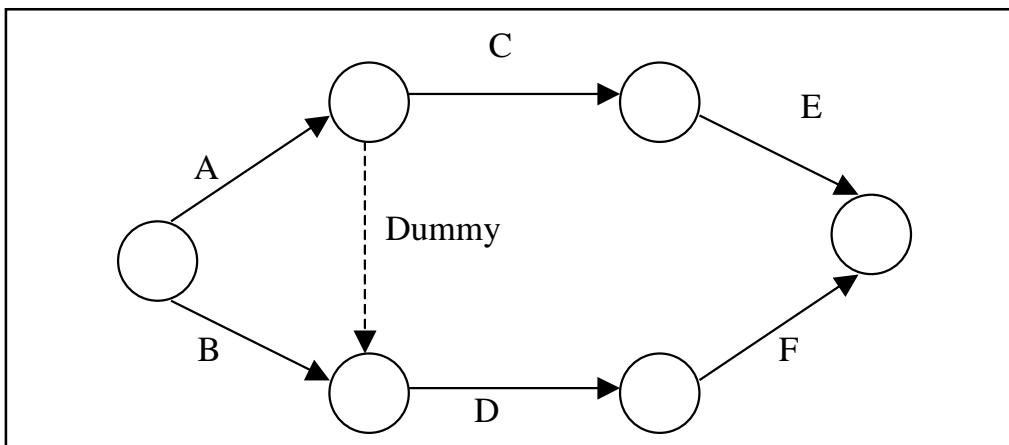


Figure 8.13: Correct representation of Network using Dummy Activity

Example 3: Construct a network for a project whose activities and their predecessor relationship are given in Table 8.3.

Table 8.3: Activity Sequence for a Project

Activity	A	B	C	D	E	F	G	H	I	J	K
Predecessor	-	-	-	A	B	B	C	D	E	H, I	F, G

Solution: The network diagram for the given problem is shown in Figure 8.14 with activities A, B and C starting simultaneously.

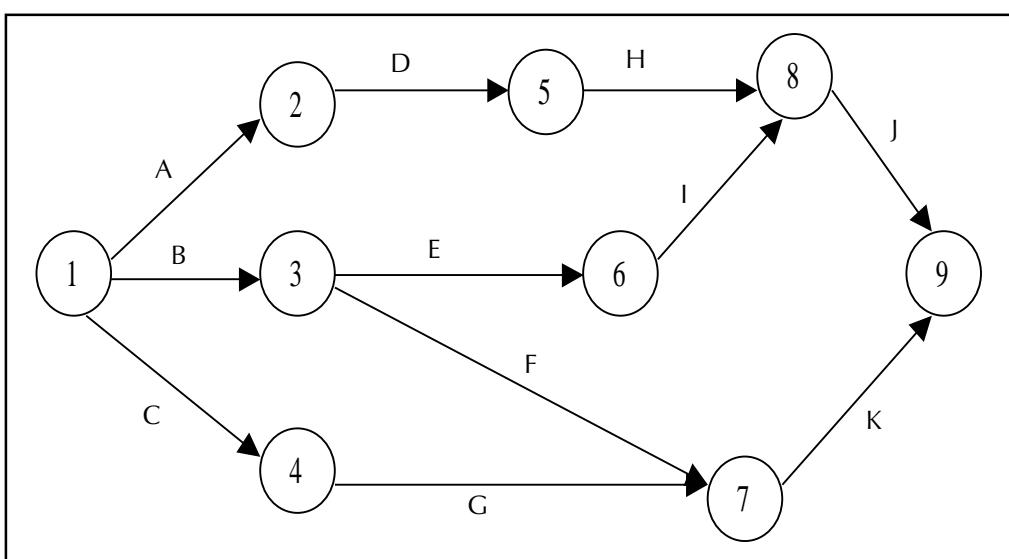


Figure 8.14: Network Diagram

Example 4: Draw a network diagram for a project given in Table 8.4.

Table 8.4: Project Activity Sequence

Activity	A	B	C	D	E	F	G	H	I	J	K	L
Immediate Predecessor	-	A	B	A	D	C, E	D	D	H	H	F, H	G, J

Solution: An activity network diagram describing the project is shown in Figure 8.15, below:

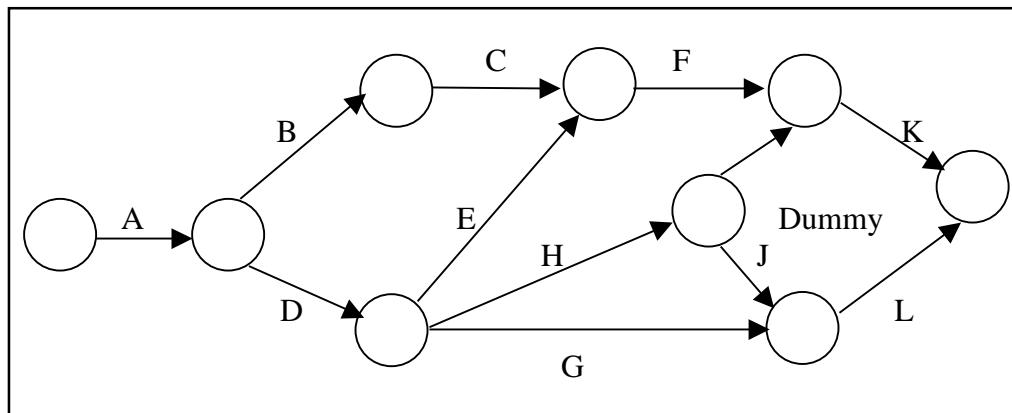


Figure 8.15: Network Diagram

8.6 CRITICAL PATH ANALYSIS

The critical path for any network is the longest path through the entire network. Since all activities must be completed to complete the entire project, the length of the critical path is also the shortest time allowable for completion of the project. Thus if the project is to be completed in that shortest time, all activities on the critical path must be started as soon as possible. These activities are called **critical activities**. If the project has to be completed ahead of the schedule, then the time required for at least one of the critical activity must be reduced. Further, any delay in completing the critical activities will increase the project duration.

The activity, which does not lie on the critical path, is called **non-critical activity**. These non-critical activities may have some slack time. The slack is the amount of time by which the start of an activity may be delayed without affecting the overall completion time of the project. But a critical activity has no slack. To reduce the overall project time, it would require more resources (at extra cost) to reduce the time taken by the critical activities to complete.

Scheduling of Activities: Earliest Time and Latest Time

Before the critical path in a network is determined, it is necessary to find the earliest and latest time of each event to know the earliest expected time (T_E) at which the activities originating from the event can be started and to know the latest allowable time (T_L) at which activities terminating at the event can be completed.

Forward Pass Computations (to calculate Earliest, Time T_E)

Procedure

Step 1: Begin from the start event and move towards the end event.

Step 2: Put $T_E = 0$ for the start event.

Step 3: Go to the next event (i.e node 2) if there is an incoming activity for event 2, add calculate T_E of previous event (i.e event 1) and activity time.

Note: If there are more than one incoming activities, calculate T_E for all incoming activities and take the maximum value. This value is the T_E for event 2.

Step 4: Repeat the same procedure from step 3 till the end event.

Procedure

Step 1: Begin from end event and move towards the start event. Assume that the direction of arrows is reversed.

Step 2: Latest Time T_L for the last event is the earliest time, T_E , of the last event.

Step 3: Go to the next event, if there is an incoming activity, subtract the value of T_L of previous event from the activity duration time. The arrived value is T_L for that event. If there are more than one incoming activities, take the minimum T_E value.

Step 4: Repeat the same procedure from step 2 till the start event.

Check Your Progress 8.1

1. What are the differences between critical and non-critical?
 2. Discuss procedural steps of Hungarian method for solving assignment problem.

Notes: (a) Write your answer in the space given below.

(b) Please go through the lesson sub-head thoroughly you will get your answers in it.

(c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

8.7 DETERMINATION OF FLOAT AND SLACK TIMES

As discussed earlier, the non – critical activities have some slack or float. The ***float*** of an activity is the amount of time available by which it is possible to delay its completion time without extending the overall project completion time.

For an activity $i = j$, let

t_{ui} = duration of activity

T_e = earliest expected time

T_L = latest allowable time

ES_{ii} = earliest start time of the activity

EF_{ij} = earliest finish time of the activity

LS_{ij} = latest start time of the activity

LF_{ii} = latest finish time of the activity

Total Float TF_{ij} : The total float of an activity is the difference between the latest start time and the earliest start time of that activity.

or

Free Float FF_{ij} : The time by which the completion of an activity can be delayed from its earliest finish time without affecting the earliest start time of the succeeding activity is called free float.

$$FF_{ij} = (E_j - E_i) - t_{ij} \quad \dots \dots \dots (3)$$

FF_{ij} = Total float – Head event slack

Independent Float IF_{ij} : The amount of time by which the start of an activity can be delayed without affecting the earliest start time of any immediately following activities, assuming that the preceding activity has finished at its latest finish time.

$$IF_{ij} = (E_j - L_i) - t_{ij} \quad \dots \dots \dots (4)$$

IF_{ij} = Free float – Tail event slack

Where tail event slack = $L_i - E_i$

The negative value of independent float is considered to be zero.

Critical Path: After determining the earliest and the latest scheduled times for various activities, the minimum time required to complete the project is calculated. In a network, among various paths, the longest path which determines the total time duration of the project is called the **critical path**. The following conditions must be satisfied in locating the critical path of a network.

An activity is said to be critical only if both the conditions are satisfied.

1. $T_L - T_E = 0$
2. $T_{Lj} - t_{ij} - T_{Ej} = 0$

Example 8.5: A project schedule has the following characteristics as shown in Table 8.5

Table 8.5: Project Schedule

Activity	Name	Time	Activity	Name	Time (days)
1-2	A	4	5-6	G	4
1-3	B	1	5-7	H	8
2-4	C	1	6-8	I	1
3-4	D	1	7-8	J	2
3-5	E	6	8-10	K	5
4-9	F	5	9-10	L	7

- i. Construct PERT network.
- ii. Compute T_E and T_L for each activity.
- iii. Find the critical path.

Solution:

- (i) From the data given in the problem, the activity network is constructed as shown in Figure 8.16.

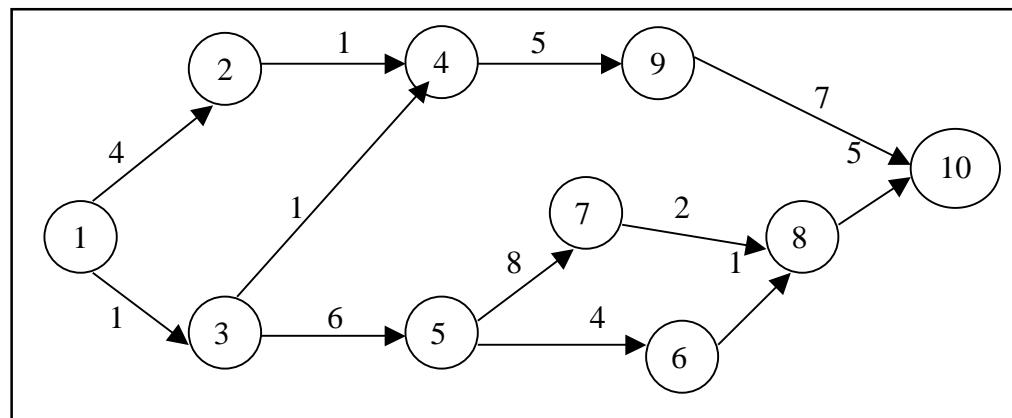


Figure 8.16: Activity Network Diagram

- (ii) To determine the critical path, compute the earliest time T_E and latest time T_L for each of the activity of the project. The calculations of T_E and T_L are as follows:

To calculate T_E for all activities,

$$\begin{aligned}
 T_{E1} &= 0 \\
 T_{E2} &= T_{E1} + t_{1,2} = 0 + 4 = 4 \\
 T_{E3} &= T_{E1} + t_{1,3} = 0 + 1 = 1 \\
 T_{E4} &= \max(T_{E2} + t_{2,4} \text{ and } T_{E3} + t_{3,4}) \\
 &= \max(4 + 1 \text{ and } 1 + 1) = \max(5, 2) \\
 &= 5 \text{ days} \\
 T_{E5} &= T_{E3} + t_{3,6} = 1 + 6 = 7 \\
 T_{E6} &= T_{E5} + t_{5,6} = 7 + 4 = 11 \\
 T_{E7} &= T_{E5} + t_{5,7} = 7 + 8 = 15 \\
 T_{E8} &= \max(T_{E6} + t_{6,8} \text{ and } T_{E7} + t_{7,8}) \\
 &= \max(11 + 1 \text{ and } 15 + 2) = \max(12, 17) \\
 &= 17 \text{ days} \\
 T_{E9} &= T_{E4} + t_{4,9} = 5 + 5 = 10 \\
 T_{E10} &= \max(T_{E9} + t_{9,10} \text{ and } T_{E8} + t_{8,10}) \\
 &= \max(10 + 7 \text{ and } 17 + 5) = \max(17, 22) \\
 &= 22 \text{ days}
 \end{aligned}$$

To calculate T_L for all activities

$$\begin{aligned}
 T_{L10} &= T_{E10} = 22 \\
 T_{L9} &= T_{E10} - t_{9,10} = 22 - 7 = 15 \\
 T_{L8} &= T_{E10} - t_{8,10} = 22 - 5 = 17 \\
 T_{L7} &= T_{E8} - t_{7,8} = 17 - 2 = 15 \\
 T_{L6} &= T_{E8} - t_{6,8} = 17 - 1 = 16 \\
 T_{L5} &= \min(T_{E6} - t_{5,6} \text{ and } T_{E7} - t_{5,7}) \\
 &= \min(16 - 4 \text{ and } 15 - 8) = \min(12, 7) \\
 &= 7 \text{ days} \\
 T_{L4} &= T_{L9} - t_{4,9} = 15 - 5 = 10 \\
 T_{L3} &= \min(T_{L4} - t_{3,4} \text{ and } T_{L5} - t_{3,5}) \\
 &= \min(10 - 1 \text{ and } 7 - 6) = \min(9, 1) \\
 &= 1 \text{ day} \\
 T_{L2} &= T_{L4} - t_{2,4} = 10 - 1 = 9 \\
 T_{L1} &= \min(T_{L2} - t_{1,2} \text{ and } T_{L3} - t_{1,3}) \\
 &= \min(9 - 4 \text{ and } 1 - 1) = 0
 \end{aligned}$$

Table 8.6: Various Activities and their Floats

Activity	Activity Name	Normal Time	Earliest Time		Latest Time		Total Float
			Start	Finish	Start	Finish	
1-2	A	4	0	4	5	9	5
1-3	B	1	0	1	0	1	0
2-4	C	1	4	5	9	10	5

Contd...

3-4	D	1	1	2	9	10	8
3-5	E	6	1	7	1	7	0
4-9	F	5	5	10	10	15	5
5-6	G	4	7	11	12	16	5
5-7	H	8	7	15	7	15	0
6-8	I	1	11	12	16	17	5
7-8	J	2	15	17	15	17	0
8-10	K	5	17	22	19	22	0
9-10	L	7	10	17	15	22	5

- (iii) From the Table 8.6, we observe that the activities 1 – 3, 3 – 5, 5 – 7, 7 – 8 and 8 – 10 are critical activities as their floats are zero.

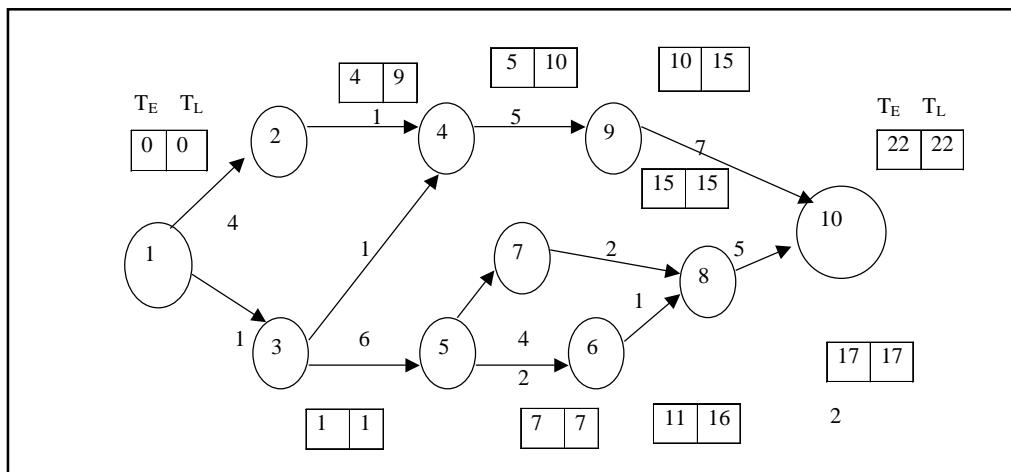


Figure 8.17: Critical Path of the Project

The critical path is 1-3-5-7-8-10 (shown in double line in Figure 8.17) with the project duration of 22 days.

Check Your Progress 8.2

Which does a critical path actually signify in a project i.e. in what ways does it differ from any other path? And What ways are its activities particularly impossible?

- Notes:**
- (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-

8.8 SOLVING CPM PROBLEMS USING COMPUTER

The worked out example, Ex. 5.5 is solved using computer.

Go to **MAIN MENU** and select **PROJECT PLANNING** and **CPM – CRITICAL PATH METHOD**. Enter the values of the network problem as shown in Figure 8.17.

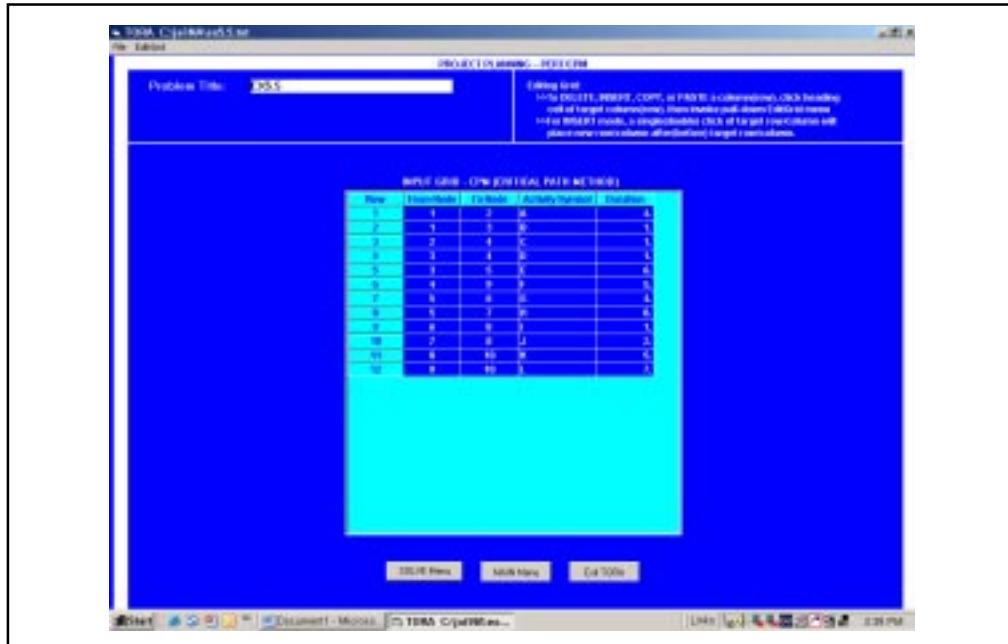


Figure 8.18: Solving Network Problem on Computer Using TORA (Input Screen)

Now select **SOLVE MENU** and **GO TO OUTPUT SCREEN**. There are two options for output, select CPM calculations. For step-by-step calculation of earliest time and latest time using forward pass and backward pass procedure click **NEXT STEP** button. To get all the values instantly, then press **ALL STEPS** button. The screen gives all the required values to analyze the problem. You may note that at the bottom of the table, the critical activities are highlighted in red colour. The output screen is shown in Figure 8.19, below:



Figure 8.19: Solving Network Problem on Computer Using TORA (Output Screen)

Example 5: The following Table 8.7 gives the activities in construction project and time duration.

Table 8.7: Project Schedule with Time Duration

Activity	Preceding Activity	Normal time (days)
1-2	-	20
1-3	-	25
2-3	1-2	10
2-4	1-2	12
3-4	1-3,2-3	5
4-5	2-4,3-4	10

- Draw the activity network of the project.
- Find the total float and free float for each activity.

Solution:

- From the activity relationship given, the activity network is shown in Figure 8.20 below:

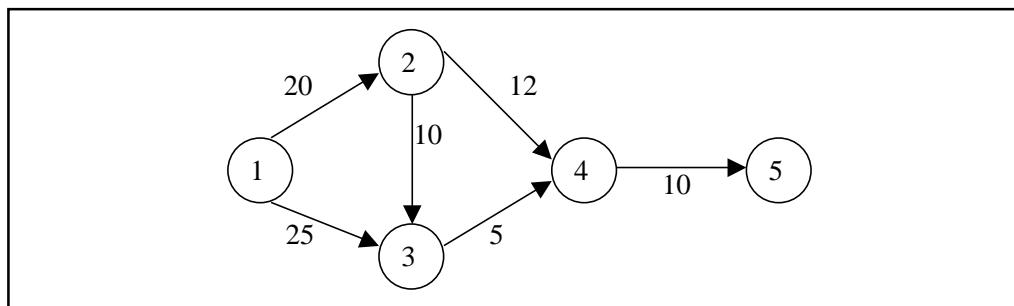


Figure 8.20: Activity Network Diagram

- The total and free floats for each activity are calculated as shown in Table 8.8

Table 8.8: Calculation of Total and Free Floats

Activity	Normal time (days)	Earliest Time		Latest Time		Float	
		Start	Finish	Start	Finish	Total	Free
1-2	20	0	20	0	20	0	0
1-3	25	0	25	5	30	5	5
2-3	10	20	30	20	30	0	0
2-4	12	20	32	23	35	3	3
3-4	5	30	35	30	35	0	0
4-5	10	35	45	35	45	0	0

Example 6: Draw the network for the following project given in Table 8.9.

Table 8.9: Project Schedule

Activity	Preceded by Initial activity	Duration (weeks)
a	-	10
b	A	9
c	A	7
d	B	6
e	B	12
f	C	6
g	C	8
h	F	8
i	D	4
j	g,h	11
k	E	5
l	I	7

Number the events by Fulkerson's rule and find the critical path. Also find the time for completing the project.

Solution: The network is drawn as shown in Figure 8.21 using the data provided. Number the events using Fulkerson's rule and find the Earliest and Latest time and total float is computed for each activity to find out the critical path as given Table 8.10.

Table 8.10: T_E , T_L and TF_{ij} Calculated

Activity	Duration weeks	Earliest Time		Latest Time		Total Float
		Start	Finish	Start	Finish	
a	10	0	10	0	10	0
b	9	10	19	16	25	6
c	7	10	17	10	17	0
d	6	19	25	25	31	6
e	12	19	31	25	37	6
f	6	17	23	17	23	0
g	8	17	25	23	31	6
h	8	23	31	23	31	0
i	4	25	29	31	35	6
j	11	31	42	31	42	0
k	5	31	36	37	42	6
l	7	29	36	35	42	6

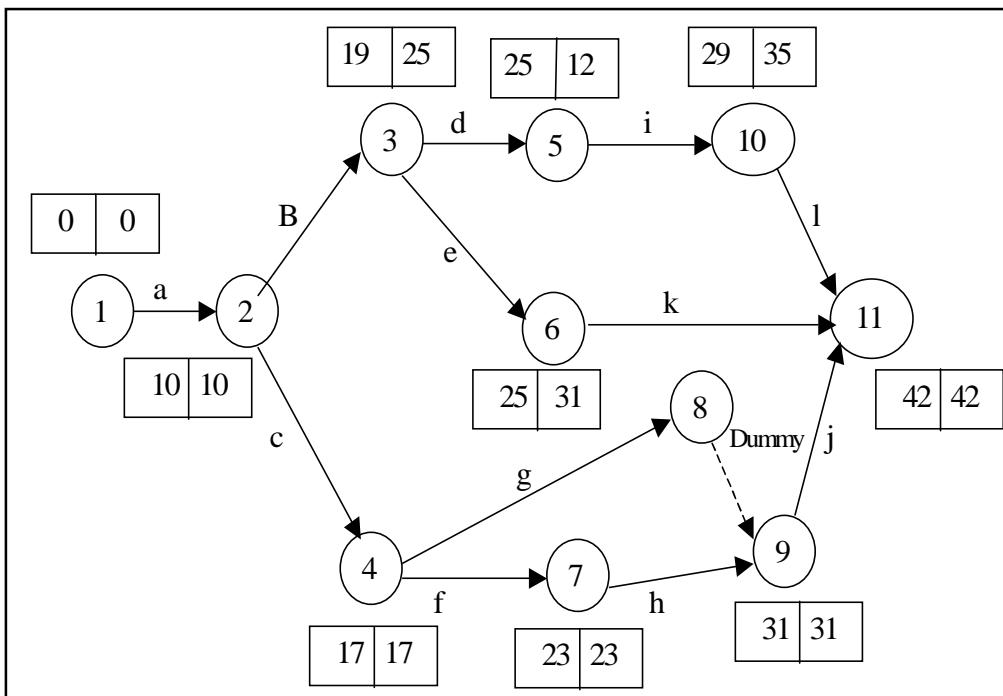


Figure 8.21: Activity Network Diagram

The critical path is $a - c - f - h - j$ and the minimum time for the completion of the project is 42 weeks.

8.9 PROJECT EVALUATION REVIEW TECHNIQUE, PERT

In the critical path method, the time estimates are assumed to be known with certainty. In certain projects like research and development, new product introductions, it is difficult to estimate the time of various activities. Hence PERT is used in such projects with a

probabilistic method using three time estimates for an activity, rather than a single estimate, as shown in Figure 8.22.

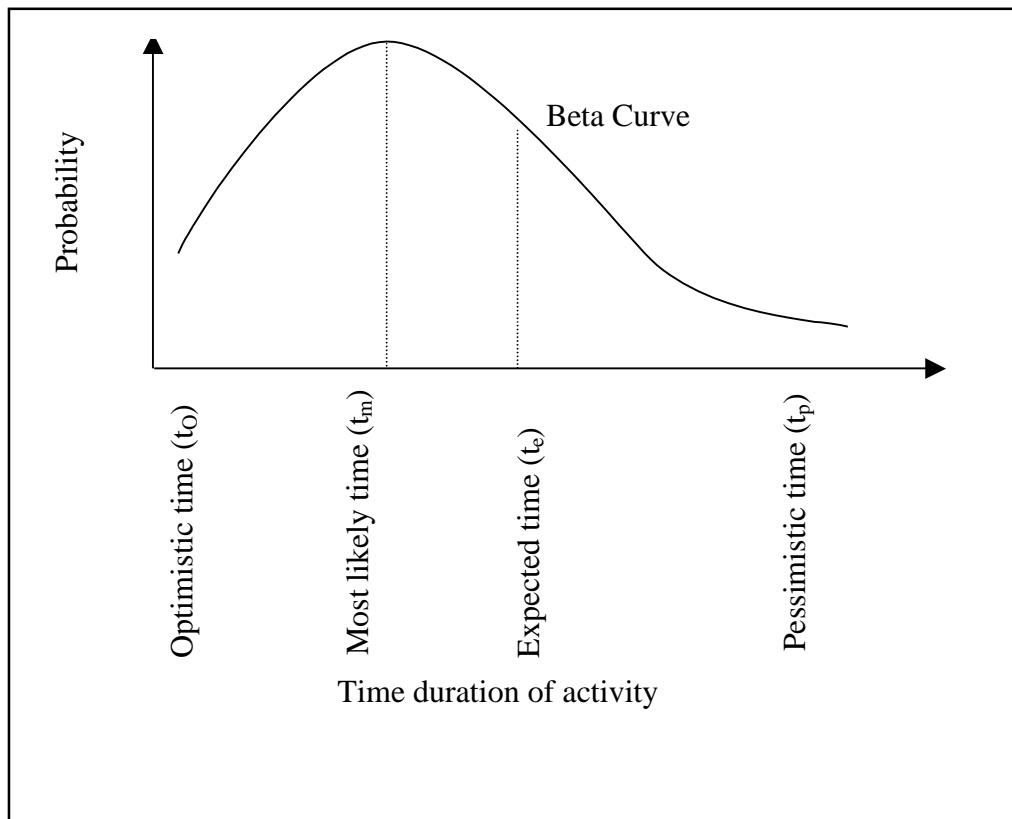


Figure 8.22: PERT Using Probabilistic Method with 3 Time Estimates

Optimistic time t_o :

It is the shortest time taken to complete the activity. It means that if everything goes well then there is more chance of completing the activity within this time.

Most likely time t_m :

It is the normal time taken to complete an activity, if the activity were frequently repeated under the same conditions.

Pessimistic time t_p :

It is the longest time that an activity would take to complete. It is the worst time estimate that an activity would take if unexpected problems are faced.

Taking all these time estimates into consideration, the expected time of an activity is arrived at.

The average or mean (t_a) value of the activity duration is given by,

$$T_a = \frac{t_0 + 4t_m + t_p}{6} \quad \dots\dots\dots(5)$$

The variance of the activity time is calculated using the formula,

$$T_a = \frac{t_0 + 4t_m + t_p}{6} \quad \dots\dots\dots(6)$$

Probability for Project Duration

The probability of completing the project within the scheduled time (T_s) or contracted time may be obtained by using the standard normal deviate where T_e is the expected time of project completion.

$$Z_0 = \frac{T_s - T_e}{\sqrt{\Sigma \sigma^2} \text{ in critical path}} \quad \dots \dots \dots (7)$$

Probability of completing the project within the scheduled time is,

$$P(T \leq T_s) = P(Z \leq Z_0) \text{ (from normal tables)} \quad \dots \quad (8)$$

Example 8: An R & D project has a list of tasks to be performed whose time estimates are given in the Table 8.11, as follows.

Time expected for each activity is calculated using the formula (5):

$$T_a = \frac{t_0 + 4tm + tp}{6}$$

$$= \frac{4 + 4(6) + 8}{6} = \frac{36}{6} = 6 \text{ days for activity A}$$

Similarly, the expected time is calculated for all the activities.

The variance of activity time is calculated using the formula (6).

$$\sigma_i^2 = \left(\frac{t_p - t_0}{6} \right)^2$$

$$= \left(\frac{8 - 4}{6} \right)^2 = 0.444$$

Similarly, variances of all the activities are calculated. Construct a network diagram and calculate the time earliest, T_E and time Latest T_L for all the activities.

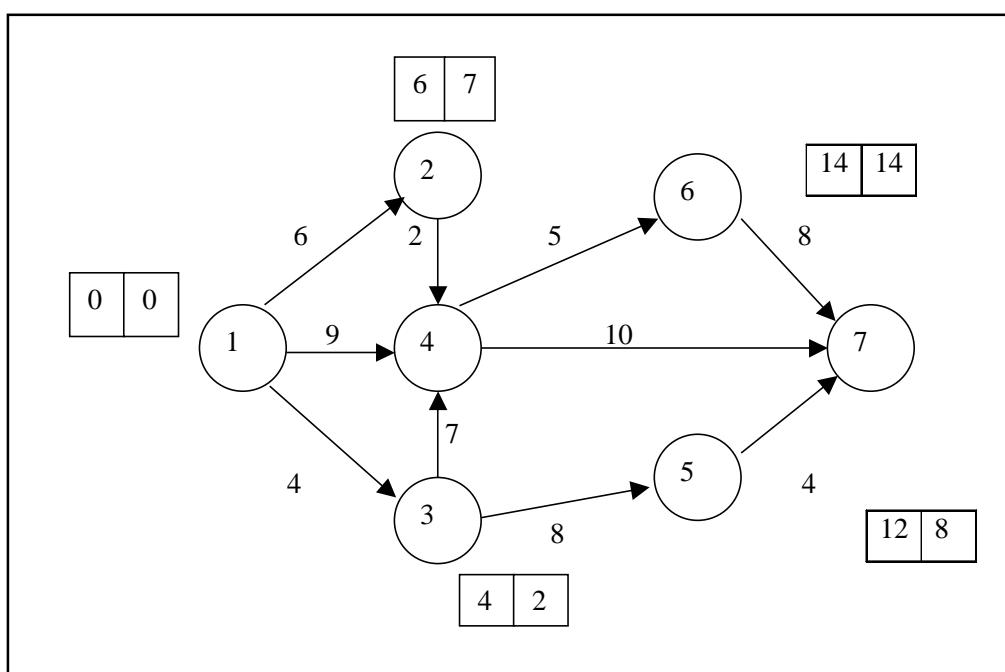


Figure 8.23: Network Diagram

Table 8.11: Time Estimates for R & D Project

Activity <i>i</i>	Activity Name	T_o	t_m (in days)	T_p
1-2	A	4	6	8
1-3	B	2	3	10
1-4	C	6	8	16
2-4	D	1	2	3
3-4	E	6	7	8
3-5	F	6	7	14
4-6	G	3	5	7
4-7	H	4	11	12
5-7	I	2	4	6
6-7	J	2	9	10

- a. Draw the project network.
- b. Find the critical path.
- c. Find the probability that the project is completed in 19 days. If the probability is less than 20%, find the probability of completing it in 24 days.

Solution:

Calculate the time average t_a and variances of each activity as shown in Table 8.12.

Table 8.12: T_e & s^2 Calculated

Activity	T_o	T_m	T_p	T_a	σ^2
1-2	4	6	8	6	0.444
1-3	2	3	10	4	1.777
1-4	6	8	16	9	2.777
2-4	1	2	3	2	0.111
3-4	6	7	8	7	0.111
3-5	6	7	14	8	1.777
4-6	3	5	7	5	0.444
4-7	4	11	12	10	1.777
5-7	2	4	6	4	0.444
6-7	2	9	10	8	1.777

From the network diagram Figure 8.24, the critical path is identified as

1-4, 4-6, 6-7, with a project duration of 22 days.

The probability of completing the project within 19 days is given by,

$$P(Z \leq Z_0)$$

To find Z_0 ,

$$Z_0 = \left(\frac{T_s - T_e}{\sqrt{\sum \sigma^2 \text{ in critical path}}} \right)$$

$$= \left(\frac{19 - 22}{\sqrt{2.777 + 0.444 + 1.777}} \right)$$

$$= \left(\frac{-3}{\sqrt{5}} \right) = -1.3416 \text{ days}$$

we know, $P(Z < Z_0) = 0.5 - Y(1.3416)$ (from normal tables, $Y(1.3416) = 0.4099$)

$$\begin{aligned} &= 0.5 - 0.4099 \\ &= 0.0901 \\ &= 9.01\% \end{aligned}$$

Network Model

Thus, the probability of completing the R & D project in 19 days is 9.01%. Since the probability of completing the project in 19 days is less than 20%, we find the probability of completing it in 24 days.

$$Z_0 = \frac{T_s - T_e}{\sqrt{\sum \sigma^2 \text{ in critical path}}}$$

$$= \left(\frac{24 - 22}{\sqrt{5}} \right) = \left(\frac{2}{\sqrt{5}} \right) = 0.8944 \text{ days}$$

$$\begin{aligned} P(Z \leq Z_0) &= 0.5 - Y(0.8944) \quad (\text{from normal tables, } Y(0.8944) = 0.3133) \\ &= 0.5 + 0.3133 \\ &= 0.8133 \\ &= 81.33\% \end{aligned}$$

8.10 SOLVING PERT PROBLEMS USING COMPUTER

Example 8.8 is solved using computer with TORA. Go to **MAIN MENU**, **SELECT PROJECT PLANNING** and Click **PERT** – Program Evaluation Review Technique.

Enter the values as shown in Figure 8.24 below.



Figure 8.24: Solving PERT Problem Using Computer with TORA (Input Screen)

Now, go to solve menu and click. In the output screen, select Activity mean / Variance option in select output option. The following screen appears as shown in Figure 8.25.

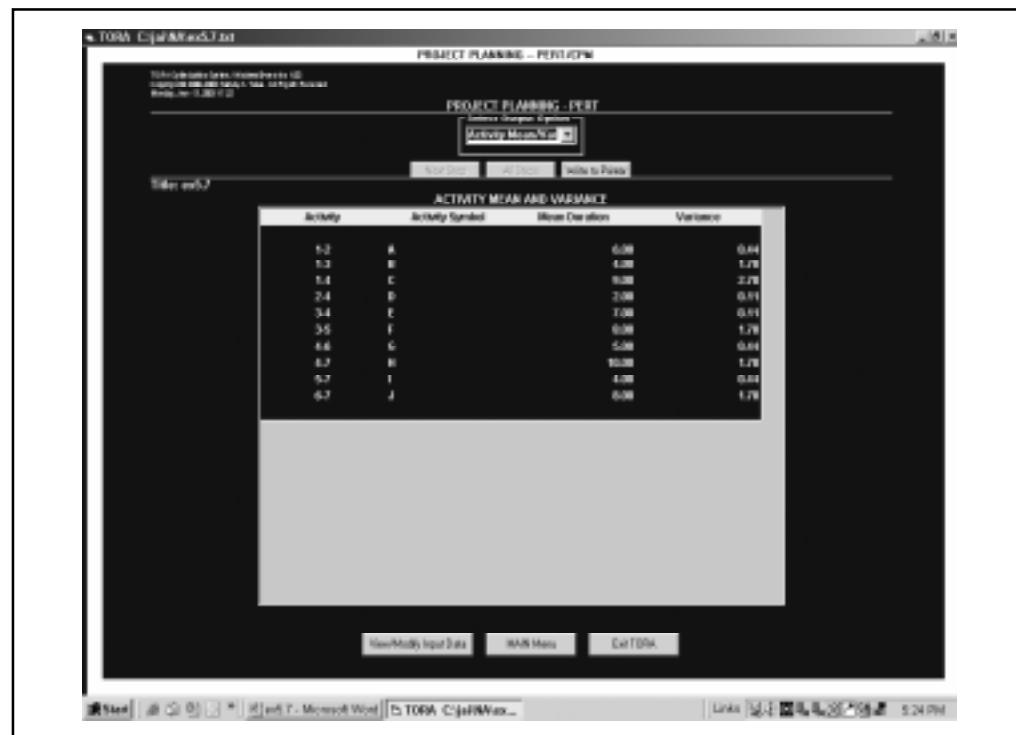


Figure 8.25: TORA (Output Screen), Select PERT Calculations.

Selecting the PERT calculations option. The following screen appears. This shows the average duration and standard deviation for the activities.



Figure 8.26: TORA (Output Screen) Showing Average Durations and Standard Deviation for Activities

8.11 COST ANALYSIS

The two important components of any activity are the cost and time. Cost is directly proportional to time and vice versa. For example, in constructing a shopping complex, the expected time of completion can be calculated using be time estimates of various

activities. But if the construction has to be finished earlier, it requires additional cost to complete the project. We need to arrive at a time / cost trade-off between total cost of project and total time required to complete it.

Normal time: Normal time is the time required to complete the activity at normal conditions and cost.

Crash time: Crash time is the shortest possible activity time; crashing more than the normal time will increase the direct cost.

Cost Slope

Cost slope is the increase in cost per unit of time saved by crashing. A linear cost curve is shown in Figure 8.27.

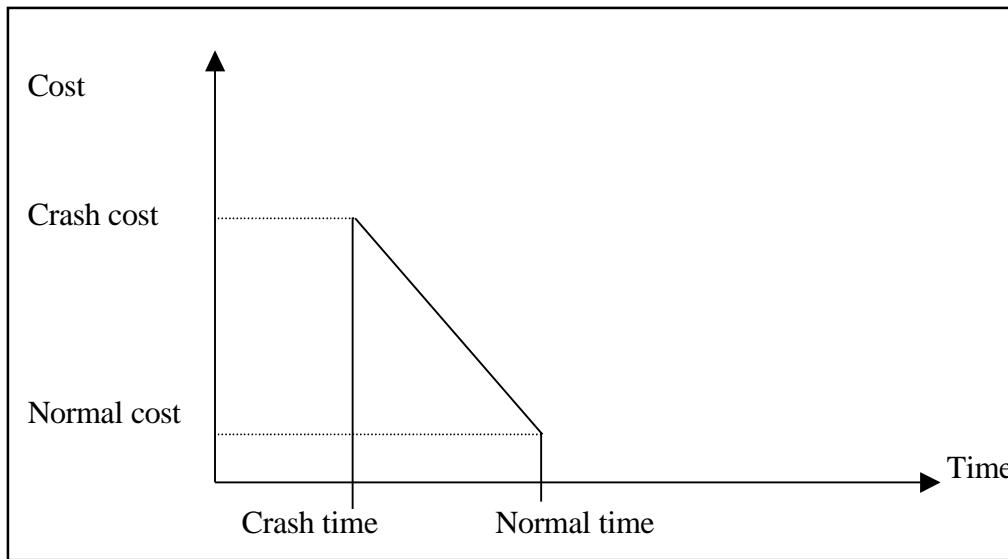


Figure 8.27: Linear Cost Curve

$$\begin{aligned} \text{Cost slope} &= \frac{\text{Crash cost } C_c - \text{Normal cost } N_c}{\text{Normal time } N_t - \text{Crash time } C_t} \\ &= \frac{C_c - N_c}{N_t - C_t} \quad \dots \dots \dots (9) \end{aligned}$$

Example 8: An activity takes 4 days to complete at a normal cost of Rs. 500.00. If it is possible to complete the activity in 2 days with an additional cost of Rs. 700.00, what is the incremental cost of the activity?

Solution:

$$\begin{aligned} \text{Incremental Cost or Cost Slope} &= \frac{C_c - N_c}{N_t - C_t} \\ &= \frac{700 - 500}{4 - 2} = \text{Rs. } 100.00 \end{aligned}$$

It means, if one day is reduced we have to spend Rs. 100/- extra per day.

Project Crashing

Procedure for crashing

Step1: Draw the network diagram and mark the Normal time and Crash time.

Step2: Calculate T_E and T_L for all the activities.

- Step 3:** Find the critical path and other paths.
Step 4: Find the slope for all activities and rank them in ascending order.
Step 5: Establish a tabular column with required field.
Step 6: Select the lowest ranked activity; check whether it is a critical activity. If so, crash the activity, else go to the next highest ranked activity.

Note: The critical path must remain critical while crashing.

- Step 7:** Calculate the total cost of project for each crashing.
Step 8: Repeat Step 6 until all the activities in the critical path are fully crashed.

Example 9: The following Table 8.13 gives the activities of a construction project and other data.

Table 8.13: Construction Project Data

Activity	Normal		Crash	
	Time (days)	Cost (Rs)	Time (days)	Cost (Rs)
1-2	6	50	4	80
1-3	5	80	3	150
2-4	5	60	2	90
2-5	8	100	6	300
3-4	5	140	2	200
4-5	2	60	1	80

If the indirect cost is Rs. 20 per day, crash the activities to find the minimum duration of the project and the project cost associated.

Solution: From the data provided in the table, draw the network diagram (Figure 8.28) and find the critical path.

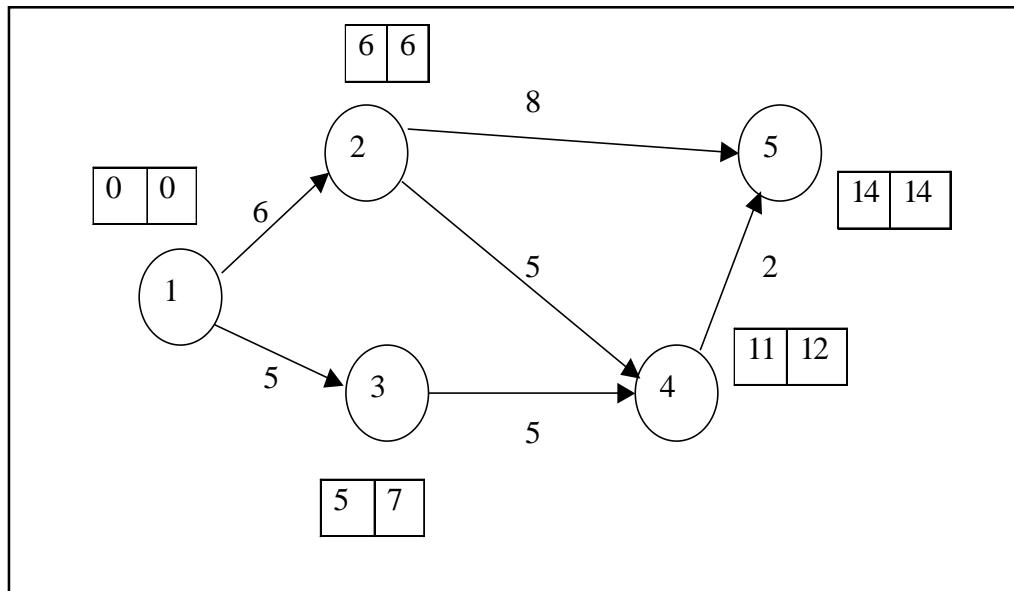


Figure 8.28: Network Diagram

From the diagram, we observe that the critical path is 1-2-5 with project duration of 14 days

The cost slope for all activities and their rank is calculated as shown in Table 8.14

$$\text{Cost slope} = \frac{\text{Crash cost } C_c - \text{Normal cost } N_c}{\text{Normal time } N_t - \text{Crash time } C_t}$$

$$\text{Cost Slope for activity } 1-2 = \frac{80 - 50}{6 - 4} = \frac{30}{2} = 15.$$

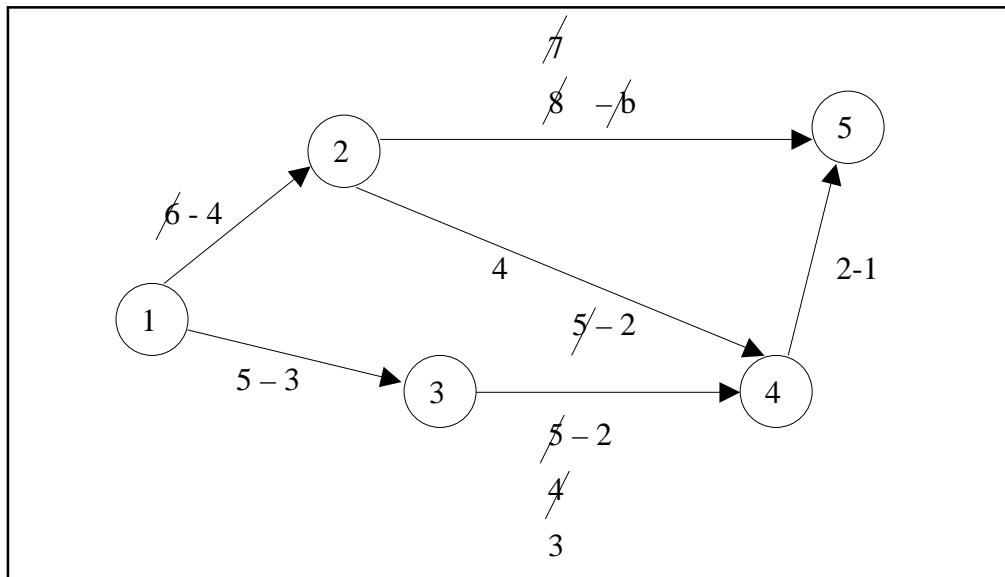
Table 8.14: Cost Slope and Rank Calculated

Activity	Cost Slope	Rank
1-2	15	2
1-3	35	4
2-4	10	1
2-5	100	5
3-4	20	3
4-5	20	3

The available paths of the network are listed down in Table 8.15 indicating the sequence of crashing (see Figure 8.29).

Table 8.15: Sequence of Crashing

Path	Number of days crashed
1-2-5	14 — 12 — 11 10
1-2-4-5	13 — 11 — 11 10
1-3-4-5	12 — 12 — 11 10

**Figure 8.29: Network Diagram Indicating Sequence of Crashing**

The sequence of crashing and the total cost involved is given in Table 8.16

Initial direct cost = sum of all normal costs given

$$= \text{Rs. } 490.00$$

Table 8.16: Sequence of Crashing & Total Cost

Activity Crashed	Project Duration	Critical Path	Direct Cost in (Rs.)	Indirect Cost (in Rs.)	Total Cost (in Rs.)
—	14	1 - 2 - 5	490	$14 \times 20 = 280$	770
1 - 2(2)	12	1 - 2 - 5	$490 + (2 \times 15) = 520$	$12 \times 20 = 240$	760

Contd...

$2 - 5 (1)$	11	$1 - 2 - 5$	$520 + (1 \times 100) + (1 \times 20) =$	$11 \times 20 = 220$	860
$3 - 4 (1)$		$1 - 3 - 4 - 5$	640		
		$1 - 2 - 4 - 5$			
$2 - 5 (1)$	10	$1 - 2 - 5$	$640 + (1 \times 100) + (1 \times 10) +$	$10 \times 20 = 200$	970
$2 - 4 (1)$		$1 - 3 - 4 - 5$	$(1 \times 20) = 770$		
$3 - 4 (1)$		$1 - 2 - 4 - 5$			

It is not possible to crash more than 10 days, as all the activities in the critical path are fully crashed. Hence the minimum project duration is 10 days with the total cost of Rs. 970.00.

Check Your Progress 8.3

If an activity zero free float, does this mean that a delay in completing that activity is likely to delay the completion of data of the project on whole.

- Notes:**
- (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Chek Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-
-

8.12 LET US SUM UP

Network analysis, as stated above, is a technique related to sequencing problems which are linked with minimizing same measure of performance of the system like the total consumption time of the project. This is a very productive technique for “describing the elements in a complex situation for purpose of designing, coordinating, planning, analysing, controlling and making decision. The new most popular form of learning’s PERT and CPM.

8.13 LESSON-END ACTIVITY

As you know that we all live in houses. Those houses are constructed by the construction company like DLF, Unitech, Parsvnath, Ansals etc. You have to visit to one of the construction company and analyse its modules-operandi to function. Apply the concept of network model (line PERT and CPM) to proper completion of work in time.

8.14 KEYWORDS

- | | |
|----------------------|---|
| Critical path | : Is a network and a continuous chain of activities that connect the initial event to the terminal event. |
| Activity | : An activity represents an action and consumption of sources. |

: Project Evaluation Review Technique — is a unique and important controlling device. The PERT take into consideration the three types of time optimistic time, pessimistic time and likely time.

: Critical Plan Method is a diagrammatical technique for planning and scheduling of projects.

: Is used in the context of network analysis. Float may be +ive or -ive.

: Direction shows the general progression in time.

: Normally associated with events. It indicates the amount of latitude.

: Is a series of related activities which result in once produces (or services). It is a pictorial presentation of the various events and activities covering a project.

: An event represent the start or completion of activity.

8.15 QUESTIONS FOR DISCUSSION

1. Write True or False against each statement:

- (a) Critical path for any network is the longest path through the entire network.
- (b) An imaginary activity always consumes resource and time.
- (c) Slack is the amount of time by which the start of an activity may be delayed.
- (d) Crash time is the maximum possible activity time.
- (e) An activity which lies on the critical path is called non-critical activity .

2. Briefly comment on the following:

- (a) PERT/CPM system results in considerable improvements.
- (b) PERT/CPM network techniques in their original developments have essentially time oriented techniques.
- (c) CPM does not incorporate statistical analysis.
- (d) Two important components of any activity are the cost and time.
- (e) Project involved planning, scheduling and controlling a number of interrelated activities.
- (f) Project managements in general have their phases.
- (g) Network should have only one start event and one end event.

3. Fill in the blank:

- (a) is the shortest possible activity time.
- (b) CPM is a time estimate and PERT has time estimate.
- (c) Cost single is the increase in cost per
- (d) time is the shortest time taken to complete the activity.

4. Write Short Notes:

- (a) PERT
- (b) CPM
- (c) Events
- (d) Activity
- (e) Crashing

8.16 TERMINAL QUESTIONS

1. What is the difference between CPM and PERT?
2. Explain the logic in constructing a network diagram. What are the network components?
3. List out the rules in constructing a network diagram.
4. What is a dummy activity?
5. What are critical path activities and why are they considered important?
6. Explain the procedure for computing earliest time and latest time of an activity.
7. What is (i) Total float (ii) Free float and (iii) Independent float ?
8. Briefly describe PERT and its advantages.
9. Explain the terms (i) Time estimates (ii) Expected time and (iii) Variance of activity time.
10. What is project crashing? Explain the procedure for crashing of project activities.

Exercise Problems

1. You are required to prepare a network diagram for constructing a 5 floor apartment. The major activities of the project are given as follows:

Activity	Description	Immediate Predecessor
A	Selection of site	-
B	Preparation of drawings	-
C	Arranging the for finance	A
D	Selection of contractor	A
E	Getting approval from Govt	A
F	Laying the foundation	E
G	Start construction	D, F
H	Advertise in newspaper	B, C
I	Allocation of tenants	G, H

2. For the problem No.1 the time estimates in days are given. Determine the Time earliest and Time latest, and the critical activities

Activity	A	B	C	D	E	F	G	H	I
Time (days)	3	5	7	2	5	20	60	2	10

3. An assembly having the following sequence of activities given along with their predecessor in the table below. Draw a network diagram for the assembly.

Activity	Description	Predecessor
A	Pick bolt & washer	-
B	Insert washer in screw	A
C	Fix the bolt in flange	A
D	Screw the nut with bolt	B, C
E	Pick the spanner	D
F	Tighten the nut	E
G	Place the assembly apart	F

4. Draw a network diagram for the project:

Activity	A	B	C	D	E	F	G	H	I	J
Predecessor	-	A	B	B	B	C	C	F, G	D, E, F	I

5. Determine the critical path and project duration for the following project:

Activity	Immediate Predecessor	Time (days)
A	-	3
B	-	7
C	A	4
D	B	2
E	C,D	5
F	A	6
G	E,F	3

6. A national conference is planned in a college. The activities are listed down along with their predecessors and time taken. Prepare a network diagram and determine the critical activities.

Activity	Description	Immediate Predecessor	Duration (days)
A	Confirm lead speaker and topic	-	5
B	Prepare brochure	-	1
C	Send letters to other speakers	B	2
D	Get confirmation from speakers	C	5
E	Send letters to participants	C,D	2
F	Obtain travel plans from speakers	D	2
G	Arrange for accommodation for speakers	F	1
H	Get handouts from speakers	F	4
I	Finalize registrations	G,H	10
J	Arrange hall and AV	I	1
K	Conduct of programme	J	1

7. Consider the following project with the list of activities:

Activity	Predecessors	Duration (months)
A	-	1
B	A	3
C	B	4
D	B	3
E	B	3
F	C	4
G	D,E	5
H	F	1
I	G,H	4

Contd....

J	I	3
K	I	4
L	J	3
M	K	5
N	L	5

- a. Construct the project network diagram.
- b. Compute the earliest start time and earliest finish time.
- c. Find the latest start and latest finish time.
- d. Find the slack for each activity.
- e. Determine the critical path and project duration.

Use TORA to compare and check answer.

8. You are alone at home and have to prepare a bread sandwich for yourself. The preparation activities and time taken are given in the table below:

Task	Description	Predecessor	Time (minutes)
A	Purchase of bread	-	20
B	Take cheese and apply on bread	A	5
C	Get onions from freezer	A	1
D	Fry onions with pepper	B,C	6
E	Purchase sauce for bread	-	15
F	Toast Bread	B,C	4
G	Assemble bread and fried onions	F	2
H	Arrange in plate	G	1

- a. Determine the critical activities and preparation time for tasks given in table.
 - b. Find the earliest time and latest time for all activities.
 - c. While purchasing sauce, you met a friend and spoke to him for 6 minutes. Did this cause any delay in preparation?
9. An amusement park is planned at a suitable location. The various activities are listed with time estimates. Using TORA, determine the critical path. Also, find whether the amusement park can be opened for public within 35 days from the start of the project work.

Activity : A B C D E F G H I J K

Time (days) : 9 6 2 7 10 3 6 1 7 2 5

The predecessor activities are given below:

Activity	Predecessor
A	-
B	-
C	A
D	A
E	C
F	B
G	E,F
H	D
I	H,E
J	I
K	G

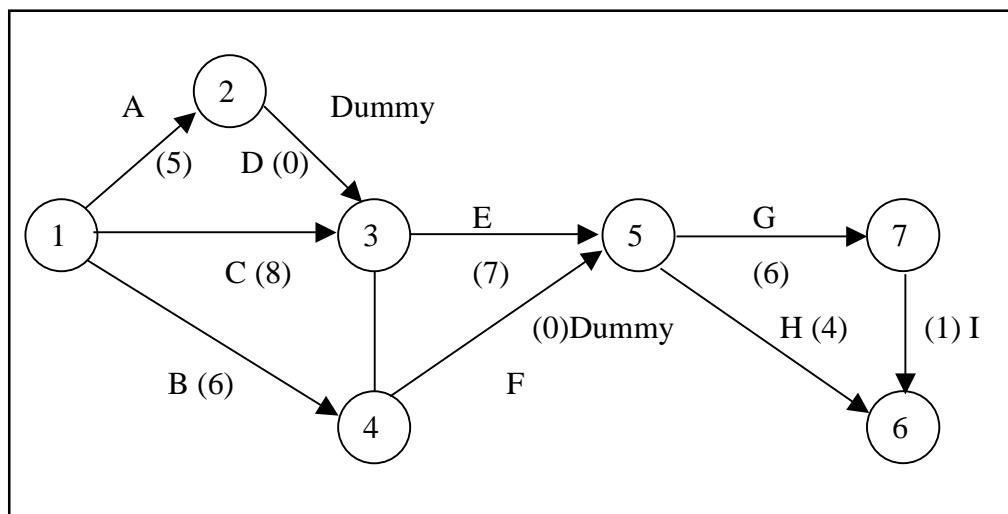
10. Draw the network from the following activity and find the critical path and total duration of project.

Activity	Duration (days)
1-2	5
1-3	3
1-4	6
2-3	8
2-5	7
3-5	2
4-5	6

11. Draw a network diagram and determine the project duration

Activity	Duration (weeks)
1-2	2
1-4	4
1-3	7
2-5	6
3-4 (Dummy)	0
4-6	6
3-6	8
5-7	10
5-6	9
5-8	2
6-7	6
7-9	2
8-9	5

12. Determine the critical path and project duration for the network given.



13. For the PERT problem find the critical path and project duration. What is the probability that the project will be completed in 25 days?

Activity	Predecessor	Time		
		Optimistic	Most likely	Pessimistic
A	-	2	5	14
B	-	1	10	12
C	A	0	0	6
D	A	1	4	7
E	C	3	10	15
F	D	3	5	7
G	B	1	2	3
H	E,F	5	10	15
I	G	3	6	9

14. The following table lists the jobs of a network along with their estimates.

Activity	Time (Weeks)		Cost (Rs)	
	Normal	Crash	Normal	Crash
1-2	9	4	1300	2400
1-3	15	13	1000	1380
2-3	7	4	7000	1540
2-4	7	3	1200	1920
2-5	12	6	1700	2240
3-6	12	11	600	700
4-5	6	2	1000	1600
5-6	9	6	900	1200

- a. Draw the project network diagram.
 - b. Calculate the length and variance of the critical path.
 - c. What is the probability that the jobs on the critical path can be completed in 41 days?
15. The following table gives data at normal time and cost crashed time and project cost.

Activity	Time (Weeks)		Cost (Rs)	
	Normal	Crash	Normal	Crash
1-2	9	4	1300	2400
1-3	15	13	1000	1380
2-3	7	4	7000	1540
2-4	7	3	1200	1920
2-5	12	6	1700	2240
3-6	12	11	600	700
4-5	6	2	1000	1600
5-6	9	6	900	1200

Network Model

Find the optimum project time and corresponding minimum total project cost by crashing appropriate activities in proper order. Show the network on time-scale at each step. Indicated cost per day is Rs. 400.00.

16. Solve the following project, and find the optimum project time and project cost.

Activity	Time (weeks)			Crash time	Cost (Rs.)	
	t ₀	t _m	t _p		Normal	Crash
1-2	1	5	3	1	500	900
2-3	1	7	4	3	800	1400
2-4	1	5	3	2	400	600
2-5	5	11	8	7	500	600
3-6	2	6	4	2	300	500
4-6	5	7	6	4	200	360
5-7	4	6	5	4	1000	1400
6-7	1	5	3	1	700	1060

8.17 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

1. (a) True (b) False (c) True (d) False (e) False
3. (a) Crash time (b) single, three (c) unit (d) optimistic

8.18 SUGGESTED READINGS

Harry & Evartis, *Introduction to PERT*

S.K. Bhatnagar, *Network Analysis Technique*

LESSON

9

WAITING MODEL (QUEUING THEORY)

CONTENTS

- 9.0 Aims and Objectives
- 9.1 Introduction
- 9.2 Queuing Systems
- 9.3 Characteristics of Queuing System
 - 9.3.1 The Arrival Pattern
 - 9.3.2 The Service Mechanism
 - 9.3.3 The Queue Discipline
 - 9.3.4 The Number of Customers allowed in the System
 - 9.3.5 The Number of Service Channels
 - 9.3.6 Attitude of Customers
- 9.4 Poisson and Exponential Distribution
- 9.5 Symbols and Notations
- 9.6 Single Server Queuing Model
- 9.7 Solving the Problem Using Computer with TORA
- 9.8 Let us Sum Up
- 9.9 Lesson-end Activity
- 9.10 Keywords
- 9.11 Questions for Discussion
- 9.12 Terminal Questions
- 9.13 Model Answers to Questions for Discussion
- 9.14 Suggested Readings

9.0 AIMS AND OBJECTIVES

In this lesson we are going to talk about the queuing theory which is also known as waiting line. These queuing theory will facilitate in solving the queue related problem of the industry. The most important point will be taken into consideration in the designing queue system which should balance service to customers.

9.1 INTRODUCTION

Queuing theory deals with problems that involve waiting (or queuing). It is quite common that instances of queue occurs everyday in our daily life. Examples of queues or long waiting lines might be

- Waiting for service in banks and at reservation counters.
- Waiting for a train or a bus.

- Waiting for checking out at the Supermarket.
- Waiting at the telephone booth or a barber's saloon.

Whenever a customer arrives at a service facility, some of them usually have to wait before they receive the desired service. This forms a queue or waiting line and customers feel discomfort either mentally or physically because of long waiting queue.

We infer that queues form because the service facilities are inadequate. If service facilities are increased, then the question arises how much to increase? For example, how many buses would be needed to avoid queues? How many reservation counters would be needed to reduce the queue? Increase in number of buses and reservation counters requires additional resource. At the same time, costs due to customer dissatisfaction must also be considered.

In designing a queuing system, the system should balance service to customers (short queue) and also the economic considerations (not too many servers). Queuing theory explores and measures the performance in a queuing situation such as average number of customers waiting in the queue, average waiting time of a customer and average server utilization.

9.2 QUEUING SYSTEMS

The customers arrive at service counter (single or in groups) and are attended by one or more servers. A customer served leaves the system after getting the service. In general, a queuing system comprises with two components, the queue and the service facility. The queue is where the customers are waiting to be served. The service facility is customers being served and the individual service stations. A general queuing system with parallel server is shown in Figure 9.1 below:

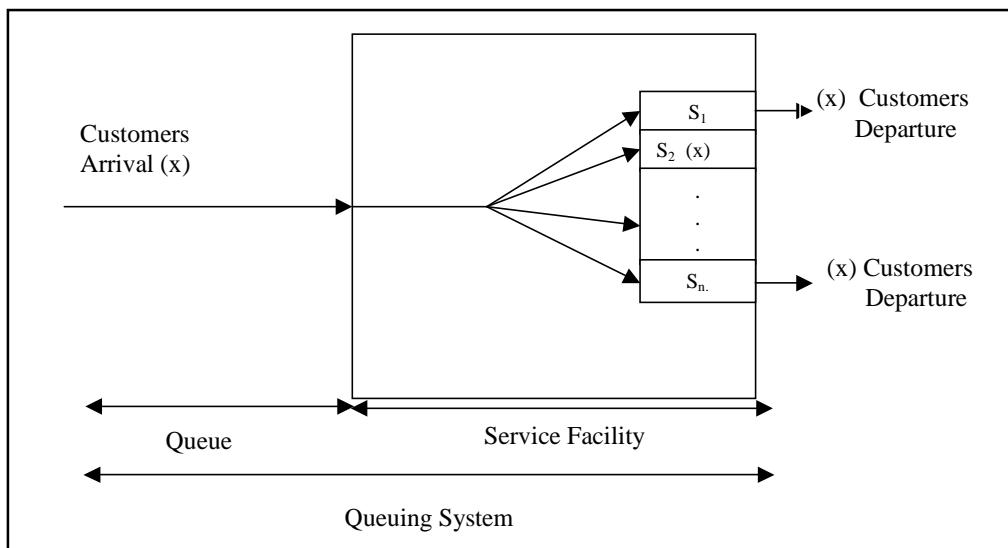


Figure 9.1: A typical queuing system

9.3 CHARACTERISTICS OF QUEUING SYSTEM

In designing a good queuing system, it is necessary to have a good information about the model. The characteristics listed below would provide sufficient information.

1. The arrival pattern.
2. The service mechanism.
3. The queue discipline.
4. The number of customers allowed in the system.
5. The number of service channels.

9.3.1 The Arrival Pattern

The arrival pattern describes how a customer may become a part of the queuing system. The arrival time for any customer is unpredictable. Therefore, the arrival time and the number of customers arriving at any specified time intervals are usually random variables. A Poisson distribution of arrivals correspond to arrivals at random. In Poisson distribution, successive customers arrive after intervals which independently are and exponentially distributed. The Poisson distribution is important, as it is a suitable mathematical model of many practical queuing systems as described by the parameter "the average arrival rate".

9.3.2 The Service Mechanism

The service mechanism is a description of resources required for service. If there are infinite number of servers, then there will be no queue. If the number of servers is finite, then the customers are served according to a specific order. The time taken to serve a particular customer is called the service time. The service time is a statistical variable and can be studied either as the number of services completed in a given period of time or the completion period of a service.

9.3.3 The Queue Discipline

The most common queue discipline is the "**First Come First Served**" (**FCFS**) or "**First-in, First-out**" (**FIFO**). Situations like waiting for a haircut, ticket-booking counters follow FCFS discipline. Other disciplines include "**Last In First Out**" (**LIFO**) where last customer is serviced first, "**Service In Random Order**" (**SIRO**) in which the customers are serviced randomly irrespective of their arrivals. "**Priority service**" is when the customers are grouped in priority classes based on urgency. "**Preemptive Priority**" is the highest priority given to the customer who enters into the service, immediately, even if a customer with lower priority is in service. "**Non-preemptive priority**" is where the customer goes ahead in the queue, but will be served only after the completion of the current service.

9.3.4 The Number of Customers allowed in the System

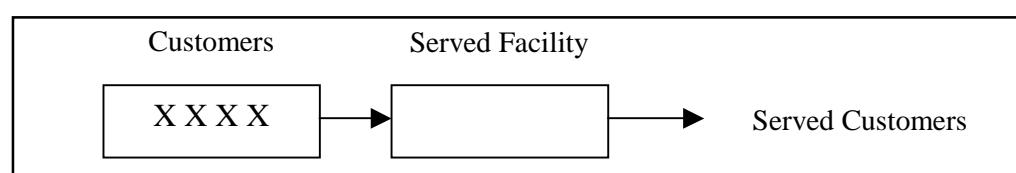
Some of the queuing processes allow the limitation to the capacity or size of the waiting room, so that the waiting line reaches a certain length, no additional customers is allowed to enter until space becomes available by a service completion. This type of situation means that there is a finite limit to the maximum queue size.

9.3.5 The Number of Service Channels

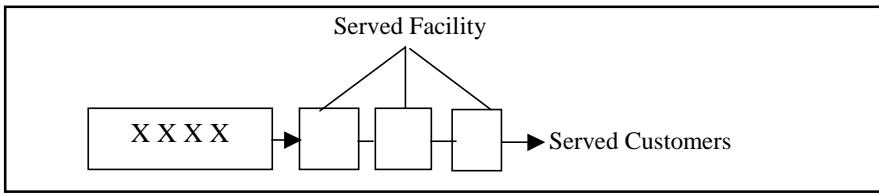
The more the number of service channels in the service facility, the greater the overall service rate of the facility. The combination of arrival rate and service rate is critical for determining the number of service channels. When there are a number of service channels available for service, then the arrangement of service depends upon the design of the system's service mechanism.

Parallel channels means, a number of channels providing identical service facilities so that several customers may be served simultaneously. Series channel means a customer go through successive ordered channels before service is completed. The arrangements of service facilities are illustrated in Figure 45. A queuing system is called a **one-server model**, i.e., when the system has only one server, and a **multi-server model** i.e., when the system has a number of parallel channels, each with one server.

(a) Arrangement of service facilities in series

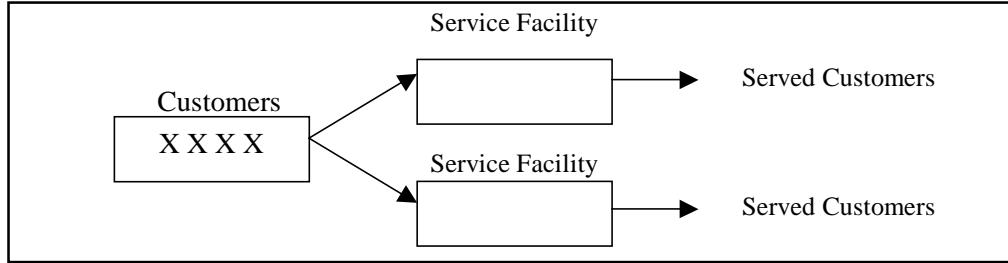


(1) Single Queue Single Server



(2) Single Queue, Multiple Server

(b) Arrangement of Service facilities in Parallel



(c) Arrangement of Mixed Service facilities

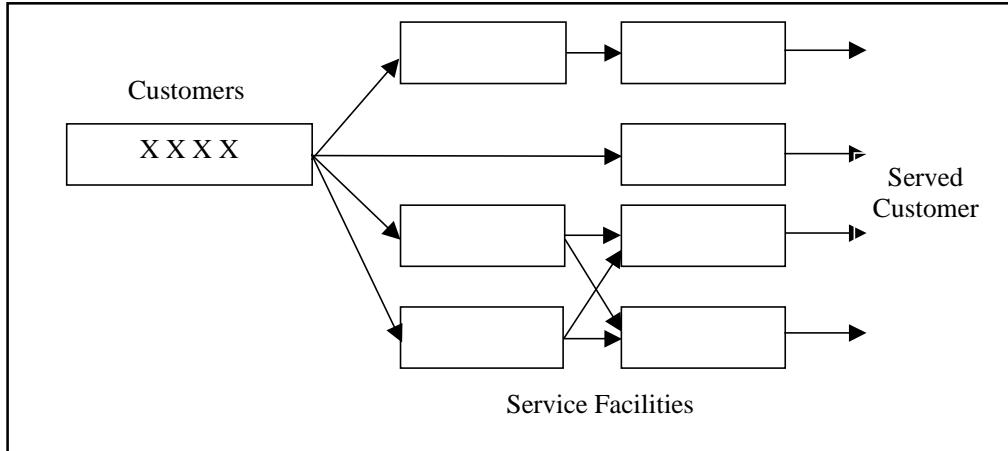


Figure 9.2: Arrangements of Service Facilities (a, b, c)

9.3.6 Attitude of Customers

Patient Customer: Customer arrives at the service system, stays in the queue until served, no matter how much he has to wait for service.

Impatient Customer: Customer arrives at the service system, waits for a certain time in the queue and leaves the system without getting service due to some reasons like long queue before him.

Balking: Customer decides not to join the queue by seeing the number of customers already in service system.

Reneging: Customer after joining the queue, waits for some time and leaves the service system due to delay in service.

Jockeying: Customer moves from one queue to another thinking that he will get served faster by doing so.

9.4 POISSON AND EXPONENTIAL DISTRIBUTIONS

Both the Poisson and Exponential distributions play a prominent role in queuing theory.

Considering a problem of determining the probability of n arrivals being observed during a time interval of length t , where the following assumptions are made.

- i. Probability that an arrival is observed during a small time interval (say of length v) is proportional to the length of interval. Let the proportionality constant be l , so that the probability is lv .
 - ii. Probability of two or more arrivals in such a small interval is zero.
 - iii. Number of arrivals in any time interval is independent of the number in non-overlapping time interval.

These assumptions may be combined to yield what probability distributions are likely to be, under Poisson distribution with exactly n customers in the system.

Suppose function P is defined as follows:

$P(n \text{ customers during period } t)$ = the probability that n arrivals will be observed in a time interval of length t

This is the Poisson probability distribution for the discrete random variable n , the number of arrivals, where the length of time interval, t is assumed to be given. This situation in queuing theory is called **Poisson arrivals**. Since the arrivals alone are considered (not departures), it is called a **pure birth process**.

The time between successive arrivals is called **inter-arrival time**. In the case where the number of arrivals in a given time interval has Poisson distribution, inter-arrival times can be shown to have the exponential distribution. If the inter-arrival times are independent random variables, they must follow an exponential distribution with density $f(t)$ where,

Thus for Poisson arrivals at the constant rate 1 per unit, the time between successive arrivals (inter-arrival time) has the exponential distribution. The average Inter - arrival time is denoted by \bar{t} .

By integration, it can be shown that $E(t) = \frac{1}{\lambda}$ (3)

If the arrival rate $\lambda = 30/\text{hour}$, the average time between two successive arrivals are $1/30 \text{ hour}$ or 2 minutes .

For example, in the following arrival situations, the average arrival rate per hour, λ and the average inter arrival time in hour, are determined.

(i) One arrival comes every 15 minutes.

$$\text{Average arrival rate , } l = \frac{60}{15} = 4 \text{ arrivals per hour.}$$

Average inter arrival time $\bar{t} = 15$ minutes $= \frac{1}{4}$ or 0.25 hour.

(ii) Three arrivals occur every 6 minutes.

Average arrival rate, $\lambda = 30$ arrivals per hour.

$$\text{Average Inter-arrival time, } \bar{I} = \frac{6}{3} = 2 \text{ minutes} = \frac{1}{30} \text{ or } 0.33 \text{ hr.}$$

(iii) Average interval between successive intervals is 0.2 hour.

Average arrival rate, $\lambda = \frac{1}{0.2} = 5$ arrivals per hour.

Average Inter-arrival time, $\bar{t} = 0.2$ hour.

Similarly, in the following service situations, the average service rate per hour, μ and average service time in hours are determined.

- (i) One service is completed in 10 minutes.

$$\text{Average service rate, } m = \frac{60}{10} = 6 \text{ services per hour.}$$

$$\text{Average service time, } \bar{s} = \frac{30}{4} = 10 \text{ minutes or } 0.166 \text{ hour.}$$

- (ii) Number of customers served in 15 minutes is 4.

$$\text{Average service rate, } m = \frac{4}{15} \times 60 = 16 \text{ services per hour.}$$

$$\text{Average service time, } \bar{s} = \frac{30}{4} = 3.75 \text{ mins or } 0.0625 \text{ hour.}$$

- (iii) Average service time is 0.25 hour.

$$\text{Average service rate, } m = 4 \text{ services per hour.}$$

$$\text{Average service time, } \bar{s} = 15 \text{ mins or } 0.25 \text{ hour.}$$

Example 1: In a factory, the machines break down and require service according to a Poisson distribution at the average of four per day. What is the probability that exactly six machines break down in two days?

Solution: Given $\lambda = 4$, $n = 6$, $t = 2$

$$P(n, t) = P(6, 4) \text{ when } t = 2$$

$$\text{we know, } P(n, t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

$$P(6, 2) = \frac{(4 \times 2)^6 e^{-4 \times 2}}{6!}$$

$$= \frac{8^6 e^{-8}}{720}$$

$$= 0.1221$$

Solving the Problem using Computer

Example 1 is solved using computer with TORA. Enter into TORA package and select Queuing Analysis option. Press 'go to input screen' to enter the values. The input screen is shown in Figure 9.3 given below. The numbers scenarios is 1 and the value of Lambda is $\lambda t = 4 \times 2 = 8$.

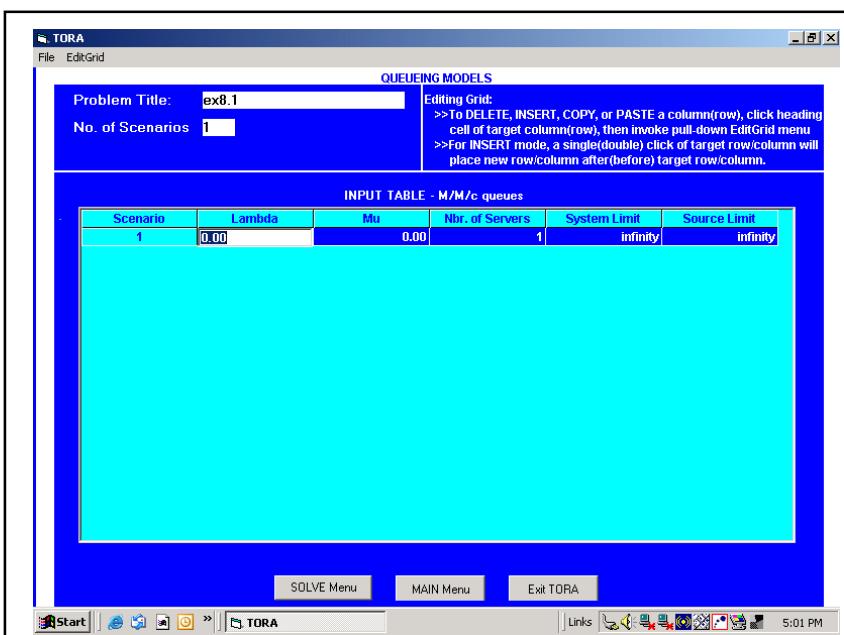


Figure 9.3: Queuing Analysis Using TORA (Input Screen)

Press 'solve', to view the Queuing Analysis output . Select Scenario 1 option, to get the result, as shown in Figure 9.4.

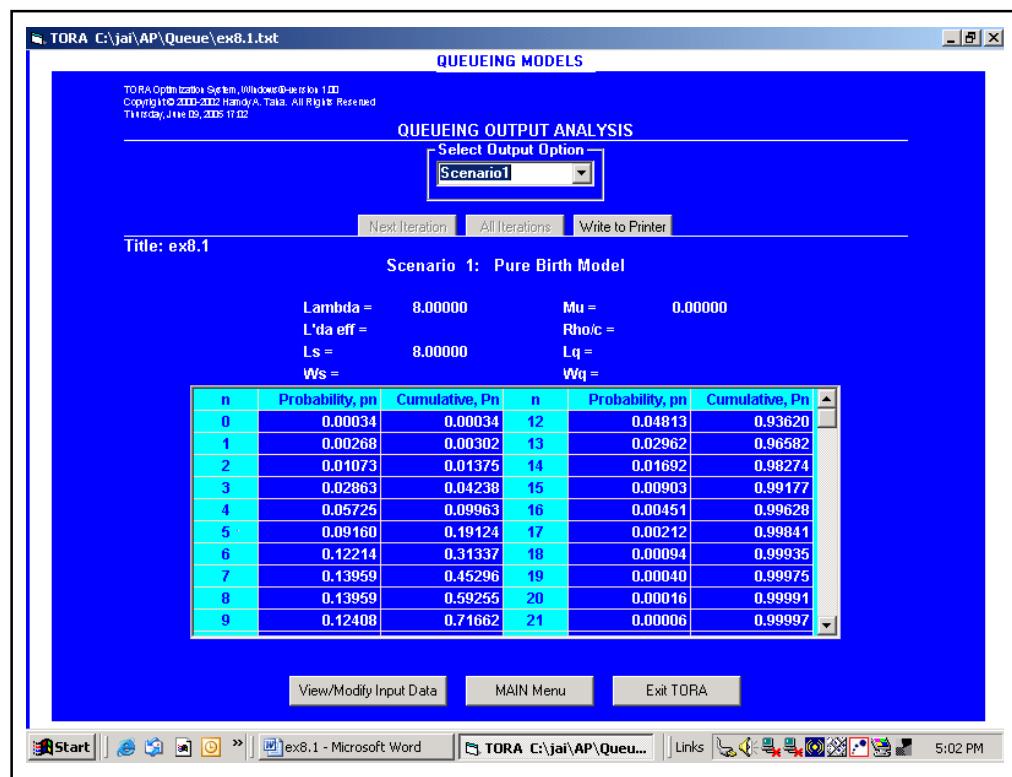


Figure 9.4: Queuing Analysis Using TORA (Output Screen)

In the output screen, for $n = 6$ the probability, P_n is given as 0.12214.

Example 2: On an average, 6 customers arrive in a coffee shop per hour. Determine the probability that exactly 3 customers will reach in a 30 minute period, assuming that the arrivals follow Poisson distribution.

Solution: Given, $\lambda = 6$ customers / hour

$$t = 30 \text{ Minutes} = 0.5 \text{ hour}$$

$$n = 2$$

we know,
$$P(n, t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

$$P(6,2) = \frac{(6 \times 0.5)^2 e^{-6 \times 0.5}}{2!} = 0.22404$$

Similarly, when the time taken to serve different customers are independent, the probability that no more than t periods would be required to serve a customer is given by exponential distribution as follows:

$$P(\text{not more than } t \text{ time period}) = 1 - e^{-\mu t} \text{ where } \mu = \text{average service rate}$$

Example 3: A manager of a fast food restaurant observes that, an average of 9 customers are served by a waiter in a one-hour time period. Assuming that the service time has an exponential distribution, what is the probability that

- (a) A customer shall be free within 12 minutes.
- (b) A customer shall be serviced in more than 25 minutes.

Solution:

(a) Given, $\mu = 9$ customers / hour

$$t = 15 \text{ minutes} = 0.25 \text{ hour}$$

Therefore, p (less than 15 minutes) = $1 - e^{-\mu t}$

$$= 1 - e^{-9 \times 0.25}$$

$$= 0.8946$$

(b) Given, $\mu = 9$ customers / hour

$$t = 25 \text{ minutes} = 0.4166 \text{ hour}$$

Therefore, P (more than 25 minutes) = $1 - e^{-\mu t}$

$$= 1 - e^{-9 \times 0.4166}$$

$$= 0.0235$$

To analyze queuing situations, the questions of interest that are typically concerned with measures of queuing system performance include,

- What will be the waiting time for a customer before service is complete?
- What will be the average length of the queue?
- What will be the probability that the queue length exceeds a certain length?
- How can a system be designed at minimum total cost?
- How many servers should be employed?
- Should priorities of the customers be considered?
- Is there sufficient waiting area for the customers?

9.5 SYMBOLS AND NOTATIONS

The symbols and notations used in queuing system are as follows:

n = Number of customers in the system (both waiting and in service).

λ = Average number of customers arriving per unit of time.

μ = Average number of customers being served per unit of time.

λ / μ = P , traffic intensity.

C = Number of parallel service channels (i.e., servers).

L_s = Average or expected number of customers in the system (both waiting and in service).

L_q = Average or expected number of customers in the queue.

W_s = Average waiting time in the system (both waiting and in service).

W_q = Average waiting time of a customer in the queue.

P_n = Time independent probability that there are n customers in the system (both waiting and in service).

$P_n(t)$ = Probability that there are n customers in the system at any time t (both waiting and in service).

9.6 SINGLE SERVER QUEUING MODEL

Model 1: (MM1) : (α / FIFO)

This model is based on the following assumptions:

- (i) The arrivals follow Poisson distribution, with a mean arrival rate λ .
 - (ii) The service time has exponential distribution, average service rate μ .
 - (iii) Arrivals are infinite population α .
 - (iv) Customers are served on a First-in, First-out basis (FIFO).
 - (v) There is only a single server.

System of Steady-state Equations

In this method, the question arises whether the service can meet the customer demand. This depends on the values of λ and μ .

If $\lambda \geq \mu$, i.e., if arrival rate is greater than or equal to the service rate, the waiting line would increase without limit. Therefore for a system to work, it is necessary that $\lambda < \mu$.

As indicated earlier, traffic intensity $\rho = \lambda / \mu$. This refers to the probability of time. The service station is busy. We can say that, the probability that the system is idle or there are no customers in the system, $P_0 = 1 - \rho$.

From this, the probability of having exactly one customer in the system is $P_1 = \rho P_0$.

Likewise, the probability of having exactly 2 customers in the system would be $P_2 = \rho P_1 = \rho^2 P_0$

The probability of having exactly n customers in the system is

$$P_n = \rho^n P_0 = \rho^{n(1-r)} = (\lambda / \mu)^n P_0$$

The expected number of customers in the system is given by,

$$L_s = \sum_n^{\alpha} n P_n = \sum_n^{\alpha} n(1 - \lambda/\mu)(\lambda/\mu)^n$$

The expected number of customers in the queue is given by,

$$L_n = \sum_{n=1}^{\alpha} (n-1)P_n$$

$$= \sum_{n=1}^{\alpha} n P_n - \sum_{n=1}^{\alpha} P_n$$

$$= \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\rho^2}{1 - \rho} \quad \dots \dots \dots (3)$$

With an average arrival rate λ , the average time between the arrivals is $1 / \lambda$. Therefore, the mean waiting time in queue, w_q is the product of the average time between the arrivals and the average queue length.

$$W_q = \left[\frac{1}{\lambda} \right] \left[\frac{1}{1-\rho} \right] \dots \dots \dots \quad (4)$$

Waiting Model
(Queuing Theory)

$$= \begin{bmatrix} 1 \\ \lambda \end{bmatrix} \begin{bmatrix} \mu \\ \mu - \lambda \end{bmatrix}$$

$$\text{Substituting } \left[\frac{\lambda^2}{\mu(\mu - \lambda)} \right] = \frac{\rho}{\mu - \lambda}$$

Similarly the average waiting time in the system , W_s

$$W_s = \left[\frac{1}{\lambda} \right] \left[\frac{\rho}{1-\rho} \right] \dots \dots \dots \quad (5)$$

putting $Ls = \lambda (m - l)$, we get

$$W_s = \frac{1}{\mu - \lambda}$$

Queuing Equations

The evaluation of Model I is listed below:

- ### 1. *Expected number of customers in the system,*

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{p}{1-p}$$

- ## 2. *Expected number of customers in the queue,*

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\rho^2}{1 - \rho}$$

- ### 3. Average waiting time in the system,

$$W_s = \frac{1}{\mu - \lambda}$$

- #### 4. Average waiting time in the queue,

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$$

- ### 5. Average waiting time for a customer,

$$W(w / w > 0) = \frac{1}{\mu(1-\rho)} \quad \text{or} \quad \frac{1}{\mu-\lambda}$$

- ### 6. *Expected length of non-empty queue,*

$$L(m/m > 0) = \frac{\mu}{(\mu - \lambda)}$$

- ### 7. Probability that there are n customers in the system,

$$P_n = \left[\frac{\lambda}{\mu} \right]^n P_0 = \left[\frac{\lambda}{\mu} \right]^n \left[1 - \frac{\lambda}{\mu} \right]$$

- ### **8. Probability that there is nobody in the system,**

$$P_0 = \frac{1-\lambda}{\mu}$$

9. **Probability that there is at least one customer or queue is busy,**

$$P_b = 1 - P_0$$

10. **Traffic intensity,**

$$\rho = \frac{\lambda}{\mu}$$

Example 4: Consider a situation where the mean arrival rate (λ) is one customer every 4 minutes and the mean service time (μ) is $2\frac{1}{2}$ minutes. Calculate the average number of customers in the system, the average queue length and the time taken by a customer in the system and the average time a customer waits before being served.

Solution: Given, Average Arrival Rate $\lambda = 1$ customer every 4 minutes or 15 customers per hour

Average Service -Rate $\mu = 1$ customer every $2\frac{1}{2}$ minutes or 24 customers per hour

- (i) **The average number of customers in the system,**

$$\begin{aligned} L_s &= \frac{\lambda}{\mu - \lambda} \\ &= \frac{15}{24 - 15} = 1.66 \text{ customers} \end{aligned}$$

- (ii) **The average queue length,**

$$\begin{aligned} L_q &= \left[\frac{\lambda}{\mu} \right] \left[\frac{\lambda}{\mu - \lambda} \right] \\ &= \frac{15}{24} \times \frac{15}{24 - 15} \\ &= 1.04 \text{ customers} \end{aligned}$$

- (iii) **The average time a customer spends in the system,**

$$\begin{aligned} W_s &= \frac{1}{\mu - \lambda} \\ &= \frac{1}{24 - 15} \\ &= 0.11 \times 60 = 6.66 \text{ minutes} \end{aligned}$$

- (iv) **The average time a customer waits before being served,**

$$\begin{aligned} W_q &= \frac{\lambda}{\lambda(\mu - \lambda)} \\ &= \frac{15}{24(24 - 15)} \\ &= 0.069 \times 60 \\ &= 4.16 \text{ minutes} \end{aligned}$$

Example 5: Trucks at a single platform weigh-bridge arrive according to Poisson probability distribution. The time required to weigh the truck follows an exponential probability distribution. The mean arrival rate is 12 trucks per day, and the mean service rate is 18 trucks per day. Determine the following:

- (a) What is the probability that no trucks are in the system?
- (b) What is the average number of trucks waiting for service?
- (c) What is the average time a truck waits for weighing service to begin?
- (d) What is the probability that an arriving truck will have to wait for service?

Solution: Given $\lambda = 12$ trucks per day, $\mu = 18$ trucks per day.

- (a) Probability that no trucks are waiting for service,

$$P_0 = 1 - \frac{\lambda}{\mu}$$

$$= 1 - \frac{12}{18}$$

$$= 0.3333 \text{ or } 33.33\%$$

- (b) Average number of trucks waiting for service,

$$L_q = \left[\frac{\lambda}{\mu} \right] \left[\frac{\lambda}{\mu - \lambda} \right]$$

$$= \left[\frac{12}{18} \right] \left[\frac{12}{18 - 12} \right]$$

$$= 1.33 \text{ trucks}$$

- (c) Average time a truck waits for weighing service to begin,

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$$

$$= \frac{12}{18(18 - 12)}$$

$$= 0.1111 \text{ days or } 53.3 \text{ minutes.}$$

- (d) Probability that an arriving truck will have to wait for service,

$$P_w = 1 - P_0$$

$$= 1 - 0.333$$

$$= 0.6667 \text{ or } 66.67\%$$

Check Your Progress 9.1

- 1 Explain Queuing Theory giving few examples.
2. “Both the Poisson and Exponential distributions play a prominent role in queuing theory.” Justify the statement.

Notes: (a) Write your answer in the space given below.

(b) Please go through the lesson sub-head thoroughly you will get your answers in it.

(c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

9.7 SOLVING THE PROBLEM USING COMPUTER WITH TORA

Example 5 is solved using computer with TORA.

Enter the values $\lambda = 12$

$$\mu = 18$$

No. of server = 1

The input screen is shown in Figure 9.5.

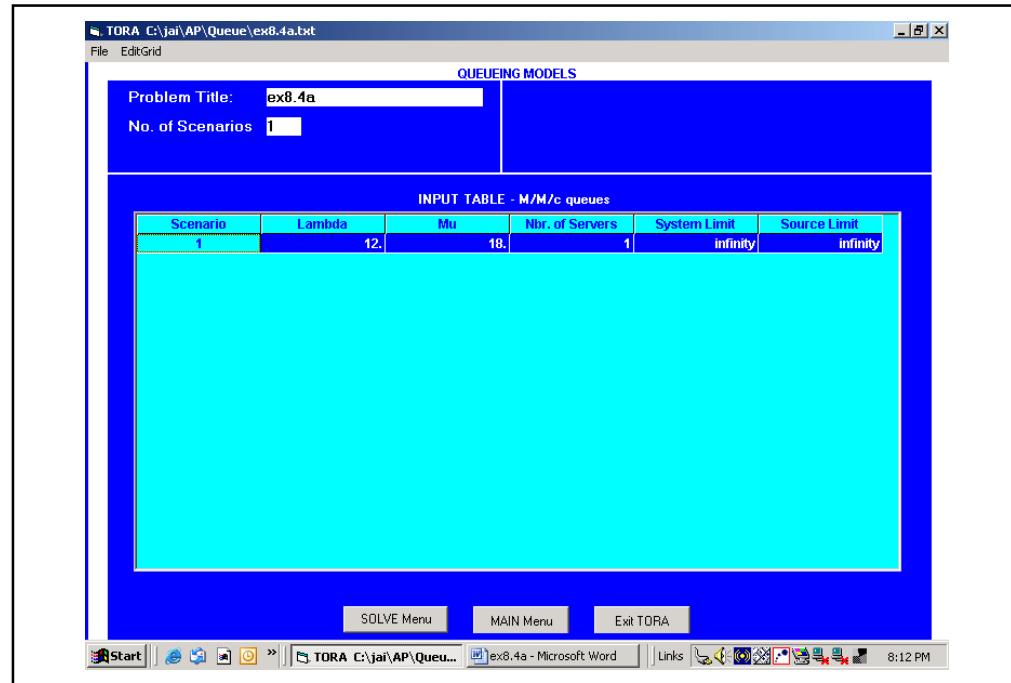


Figure 9.5: Queuing Analysis Using TORA (Input Screen)

Press Solve to get the output screen and select scenario 1 option in the select output option menu. The output screen for the problem is displayed as shown in Figure 9.6.

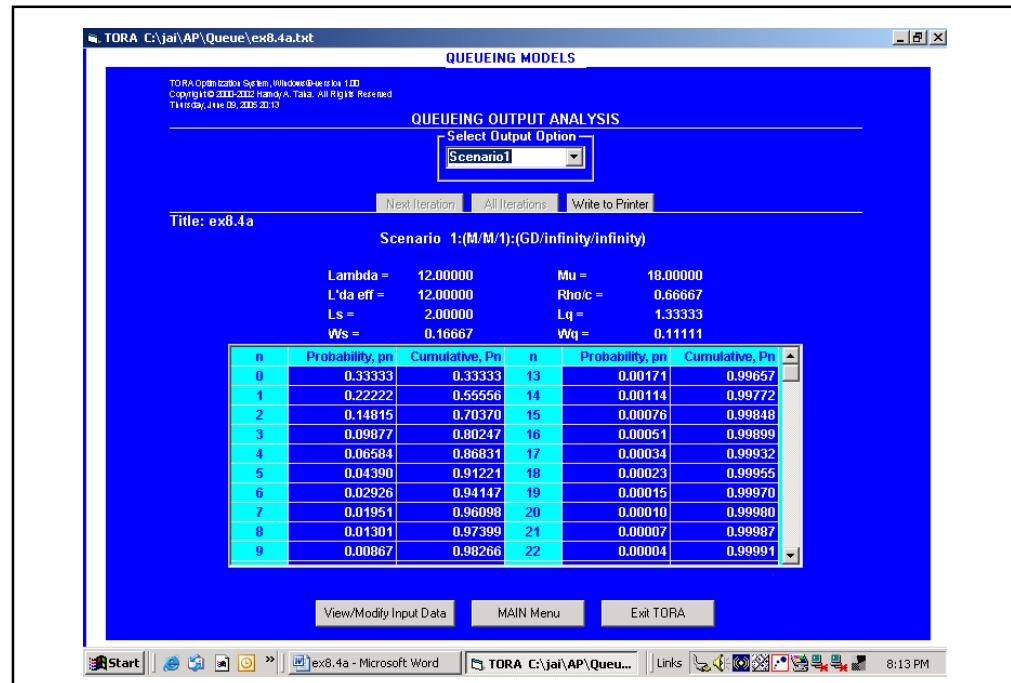


Figure 9.6: Queuing Analysis Using TORA (Output Screen)

The values are

- (a) $P_0 = 0.3333$ (for $n = 0$)
- (b) $L_q = 1.33$
- (c) $W_q = 0.1111$
- (d) P_b (or) $\frac{\rho}{C} = 0.66667$

In the same problem, to determine the probability that there are 2 trucks in the system, we use the formula,

$$\begin{aligned} P_n &= \left[\frac{\lambda}{\mu} \right]^n \left[1 - \frac{\lambda}{\mu} \right] \\ &= \left[\frac{12}{18} \right]^2 \left[1 - \frac{12}{18} \right] \\ &= 0.4444 \times 0.3333 \\ &= 0.14815 \text{ or } 14.81\% \end{aligned}$$

This can also be read in the output screen for $n=2$ the probability $P_n = 0.14815$,

Similarly, the probabilities for different values of n can be directly read.

Example 6: A TV repairman finds that the time spent on his jobs has a exponential distribution with mean 30 minutes. If he repairs TV sets in the order in which they come in, and if the arrivals follow approximately Poisson distribution with an average rate of 10 per 8 hour day, what is the repairman's expected idle time each day? How many jobs are ahead of the average with the set just brought in?

Solution: Given $\lambda = 10$ TV sets per day.

$$\mu = 16 \text{ TV sets per day.}$$

- (i) **The Probability for the repairman to be idle is,**

$$P_0 = 1 - \rho$$

$$\text{We know, } \rho = \lambda / 30 = 10 / 16 = 0.625$$

$$\begin{aligned} P_0 &= 1 - \rho \\ &= 1 - 0.625 = 0.375 \end{aligned}$$

$$\begin{aligned} \text{Expected idle time per day} &= 8 \times 0.375 \\ &= 3 \text{ hours.} \end{aligned}$$

- (ii) **How many jobs are ahead of the average set just brought in**

$$L_s = \frac{\lambda}{\mu - \lambda}$$

$$= \frac{10}{16 - 10} = \frac{10}{6}$$

$$= 1.66 \text{ say } 2 \text{ jobs.}$$

Example 7: Auto car service provides a single channel water wash service. The incoming arrivals occur at the rate of 4 cars per hour and the mean service rate is 8 cars per hour. Assume that arrivals follow a Poisson distribution and the service rate follows an exponential probability distribution. Determine the following measures of performance:

- (a) What is the average time that a car waits for water – wash to begin?
- (b) What is the average time a car spends in the system?
- (c) What is the average number of cars in the system?

Solution: Given $\lambda = 4$ cars per hour, $\mu = 8$ cars per day.

- (a) Average time that a car waits for water - wash to begin,

$$W_q = \frac{\lambda}{\lambda(\mu - \lambda)}$$

$$= \frac{4}{8(8 - 4)}$$

$$= 0.125 \text{ hours or } 7.5 \text{ mins.}$$

- (b) Average time a car spends in the system,

$$W_s = \frac{1}{\mu - \lambda}$$

$$= \frac{1}{8 - 4} = \frac{1}{4} = 0.25 \text{ hours or } 15 \text{ mins.}$$

- (c) Average number of cars in the system,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{4}{8 - 4}$$

$$= \frac{4}{4} = 1 \text{ car.}$$

Example 8: Arrivals at a telephone booth are considered to be Poisson distributed with an average time of 10 minutes between one arrival and the next. The length of phone call is assumed to be distributed exponentially, with mean 3 minutes.

- (i) What is the probability that a person arriving at the booth will have to wait?
- (ii) The telephone department will install a second booth when convinced that an arrival would expect waiting for at least 3 minutes for phone call. By how much should the flow of arrivals increase in order to justify a second booth?
- (iii) What is the average length of the queue that forms from time to time?
- (iv) What is the probability that it will take him more than 10 minutes altogether to wait for the phone and complete his call?
- (v) What is the probability that it will take him more than 10 minutes altogether to wait for the phone and complete his call?

Solution: Given $\lambda = 1/10 = 0.10$ person per minute.

$$\mu = 1/3 = 0.33 \text{ person per minute.}$$

- (i) Probability that a person arriving at the booth will have to wait,

$$P(w > 0) = 1 - P_0$$

$$= 1 - (1 - \lambda / \mu) = \lambda / \mu$$

$$= \frac{0.10}{0.33} = 0.3$$

- (ii) The installation of second booth will be justified if the arrival rate is more than the waiting time.

Expected waiting time in the queue will be,

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$$

Where, $E(w) = 3$ and $\lambda = \lambda$ (say) for second booth. Simplifying we get $\lambda = 0.16$

Hence the increase in arrival rate is, $0.16 - 0.10 = 0.06$ arrivals per minute.

- (iii) Average number of units in the system is given by,

$$L_s = \frac{\rho}{1-\rho} = \frac{0.3}{1-0.3} = 0.43 \text{ customers}$$

- (iv) Probability of waiting for 10 minutes or more is given by

$$P(W \geq 10) = \int_{10}^{\infty} \frac{\lambda}{\mu} (\mu - \lambda) e^{-(\mu - \lambda)t} dt$$

$$= \int_{10}^{\infty} (0.3) (0.23) e^{-0.23t} dt$$

$$= 0.069 \left[\frac{e^{-0.23t}}{-0.23} \right]_{10}^{\infty}$$

$$= 0.03$$

This shows that 3 percent of the arrivals on an average will have to wait for 10 minutes or more before they can use the phone.

Example 9: A bank has decided to open a single server drive-in banking facility at its main branch office. It is estimated that 20 customers arrive each hour on an average. The time required to serve a customer is 3 minutes on an average. Assume that arrivals follow a Poisson distribution and the service rate follows an exponential probability distribution.

The bank manager is interested in knowing the following:

- (a) What will be the average waiting time of a customer to get the service?
- (b) The proportion of time that the system will be idle.
- (c) The space required to accommodate all the arrivals, on an average, the space taken by each car is 10 feet that is waiting for service.

Solution: $\lambda = 20$ Customers per hour, $\mu = \frac{60}{25} = 2.4$ customers per hour.

- (a) Average waiting time of a customer to get the service,

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$$

$$= \frac{20}{24(24 - 20)} = \frac{20}{96}$$

$$= 0.208 \text{ hour or } 12.5 \text{ mins.}$$

- (b) The proportion of time that the system will be idle,

$$P_0 = 1 - \frac{\lambda}{\mu}$$

$$= 1 - \frac{20}{24}$$

$$= 0.166 \text{ hours or } 10 \text{ mins.}$$

(c) Average number of customers waiting in the queue,

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

$$= \frac{20^2}{24(24 - 20)} = \frac{400}{96}$$

$$= 4.66 \text{ customers.}$$

10 feet is required for 1 customer. Hence, for 4.66 customers, the space required is $10 \times 4.66 = 46.6$ feet.

Example 10: In a Bank, customers arrive to deposit cash to a single counter server every 15 minutes. The bank staff on an average takes 10 minutes to serve a customer. The manager of the bank noticed that on an average at least one customer was waiting at the counter. To eliminate the customer waiting time, the manager provided an automatic currency counting machine to the staff. This decreased the service time to 5 minutes on an average to every customer. Determine whether this rate of service will satisfy the manager's interest. Also use computer with TORA for solving the problem.

Solution:

Case 1: $\lambda = \frac{60}{15} = 4$ customers per hour, $\mu = \frac{60}{10} = 6$ customers per hour.

Average number of customers in the system,

$$L_s = \frac{\lambda}{\mu - \lambda}$$

$$= \frac{4}{6 - 4} = \frac{4}{2} = 2 \text{ customers.}$$

Case 2: $\lambda = 4$, $\mu = \frac{60}{15} = 12$ customers per hour.

Average number of customers in the system,

$$L_s = \frac{4}{12 - 4}$$

$$= \frac{4}{8} = \frac{1}{2} = 0.5, \text{ say, 1 customer.}$$

Average number of customers in the queue

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

$$= \frac{4^2}{12(12 - 4)} = \frac{16}{96} = 0.01 \text{ customers.}$$

Since no customers are standing in the queue the manager's interest is satisfied.

The problem is worked out using TORA. Enter the values as shown in the input screen below in Figure 9.7.

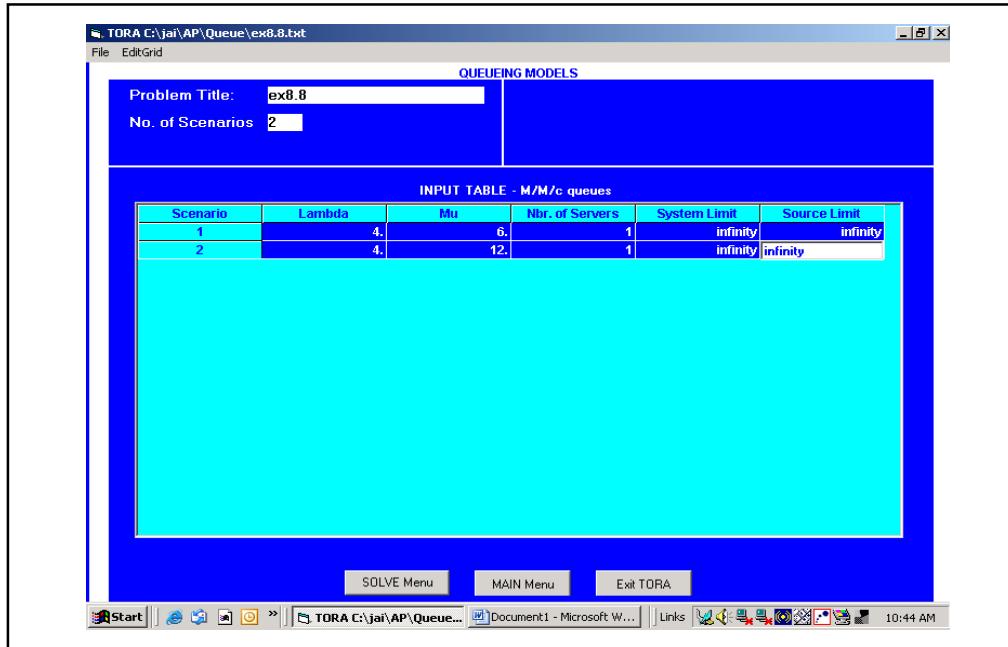


Figure 9.7: Queuing Analysis Using TORA (Input Screen)

Press Solve and go to output screen. Select comparative analysis option in the queuing output analysis menu. The following output screen is displayed (Figure 9.8).

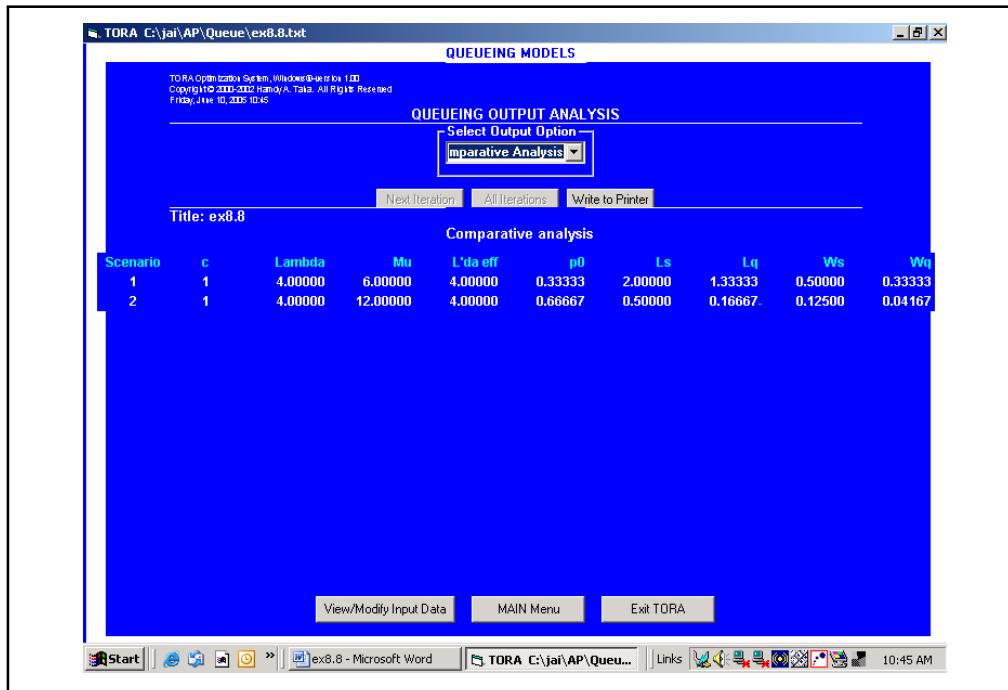


Figure 9.8: Comparative Analysis of Queuing Output Analysis Using TORA (Output Screen)

Now, on comparing scenario 1 and scenario 2, under Ls i.e., the average number of customers in the system is 2 and 0.5 respectively. In the first scenario, it means that in the entire system, one customer will be waiting in the queue while others are being served. In scenario 2, only one customer is in the system and being served, where on an average no customer will be waiting.

Example 11: 12 counters are available in a computerized railway reservation system. The arrival rate during peak hours is 90 customers per hour. It takes 5 minutes to serve a customer on an average. Assume that the arrivals joining in a queue will not be jockeying (i.e., move to another queue). How many counters have to be opened if the customers need not to wait for more than 15 minutes?

Solution: The problem is to be solved as one system comprising of 'n' number of single server queuing model.

Arrival rate, $\lambda = 90$ customers per hour

Service rate, $\mu = \frac{60}{5} = 12$ per hour

$$\text{Average waiting time, } W_q = \frac{15}{60} = 0.25 \text{ hours}$$

Average waiting time, $W_q = \frac{\lambda}{\mu(\mu - \lambda)}$

Let, number of counters be x .

Considering the single server queuing system, the number of counters required to serve

90 arrivals per hour, $\lambda = \frac{90}{x}$ substituting $\lambda = \frac{90}{x}$ in equation (i),

$$0.25 = \frac{90/x}{12\left(12 - \frac{90}{x}\right)}$$

$$0.25 = \frac{90}{12(12x - 90)}$$

$$0.25 \times 12(12x - 90) = 90$$

$$3(12x - 90) = 90$$

$$36x - 270 = 90$$

$$36x = 360$$

$$x = \frac{360}{36} = 10 \text{ counters}$$

Hence, 10 counters are required so that an average arrival will wait less than 15 minutes.

Example 12: In a single pump petrol station, vehicles arrive at the rate of 20 customers per hour and petrol filling takes 2 minutes on an average. Assume the arrival rate is Poisson probability distribution and service rate is exponentially distributed, determine

- (a) What is the probability that no vehicles are in the petrol station?
 - (b) What is the probability that 1 customer is filling and no one is waiting in the queue?
 - (c) What is the probability that 1 customer is filling and 2 customers are waiting in the queue?
 - (d) What is the probability that more than 2 customers are waiting?

Solution : $l \equiv 20$ vehicles per hour, $m \equiv 60/2 \equiv 30$ vehicles per hour.

- (a) Probability that no vehicles are in the petrol station,

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{20}{30}$$

$$= 0.3334 \text{ or } 33.34\%$$

- (b) Probability that 1 customer is filling and no one is waiting in the queue,

$$P_1 = \left[\frac{\lambda}{\mu} \right]^n P_0 = \left[\frac{\lambda}{\mu} \right]^n \left[1 - \frac{\lambda}{\mu} \right]$$

$$P_1 = \left[\frac{20}{30} \right]^1 \left[1 - \frac{20}{30} \right]$$

$$= 0.6666 \times 0.3334$$

$$= 0.2222 \text{ or } 22.22\%$$

- (c) Probability that 1 customer is filling and 2 customers are waiting in the queue, i.e., there are 3 customers in the system,

$$P_3 = \left[\frac{20}{30} \right]^3 \left[1 - \frac{20}{30} \right]$$

$$= 0.2963 \times 0.3334$$

$$= 0.09878 \text{ or } 9.87\%$$

- (d) Probability that more than 3 customers are in the system,

$$P_4 = \left[\frac{20}{30} \right]^4 \left[1 - \frac{20}{30} \right]$$

$$= 0.1975 \times 0.3334$$

$$= 0.6585 \text{ or } 65.85\%$$

The calculation made for the above problem is represented in the TORA output screen shown below in Figure 9.9.

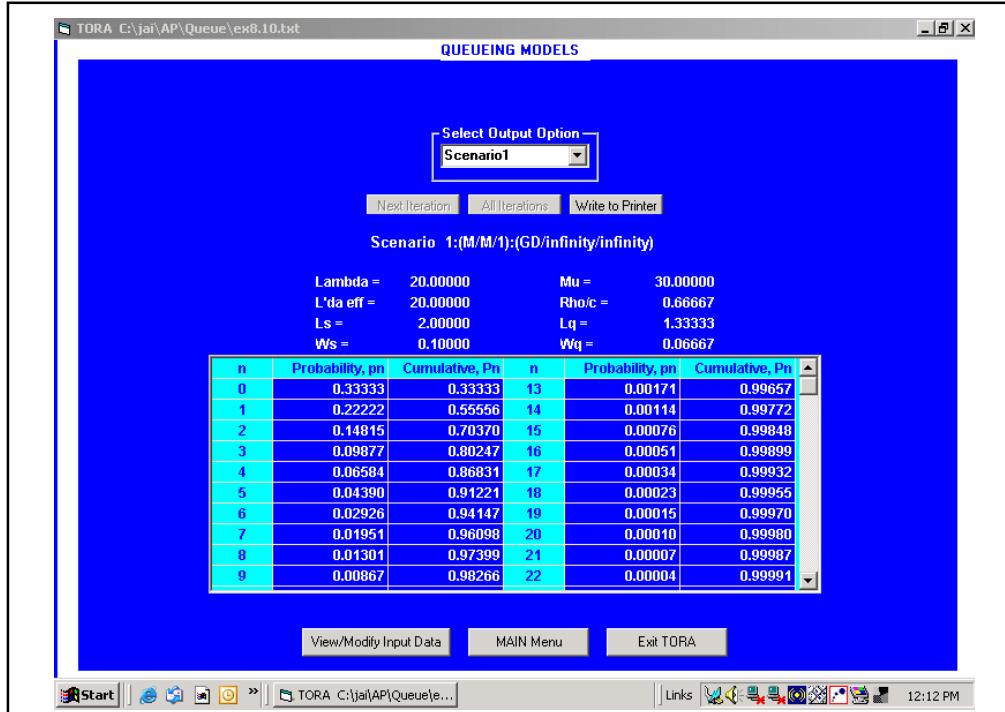


Figure 9.9: Queuing Analysis Using TORA (Output Screen)

Check Your Progress 9.2

1. The assumption of queuing theory are so restrictive as to render behaviour prediction of queuing system practically worthless Discuss.
2. Explain the meaning of a queue and state the object of queuing analysis. Briefly describe with the help of hypothetical example the elements of the queuing system.
3. Give examples of five situations/circumstances in which there in a limited a finite waiting line.
4. Elaborate the vital operating characterisits of a queuing system.
5. What are the modules of the following queuing system? Draw and explains the configuration of each
 - (a) General store
 - (b) Big Bazar
 - (c) Railway reservation
 - (d) Car wash at the service center.

Notes: (a) Write your answer in the space given below.
(b) Please go through the lesson sub-head thoroughly you will get your answers in it.
(c) This Chek Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

9.8 LET US SUM UP

Queuers or waiting lines are very familiar in day to day life. In general quite often we face the problem of long queues for a bus, a movie ticket, railway reservation, ATM m/c and various, other situation. The queuing model are those where a facility perform a service. A queuing problem arises when the current services rate of a facility fall short of the current flow rate of customer. Thus if the service facility capable of servicing customers arrive there will be no pitfalls. Thus the queuing theory is related with the decision making process of the business unit which relates with the queue question and makes decision relative to the number of service facilities which are operating.

9.9 LESSON-END ACTIVITY

As you are travelling from one place to another. You need a various mode of transportation from one destination to other. These transportation are known as Bus, Train, Aeroplane etc. While taking reservation of this particular transport. We have to go to the reservation counter and book tickets and finally face a huge waiting line or queue of passengers. Apply the waiting line theory to regulate this problem and find solution to make the system stream-line.

9.10 KEYWORDS

Balking	: A customer may not like to join the queue seeing it very long and he may not like to wait.
Reneging	: He may leave the queue due to impatience after joining in collusion several customers may collaborate and only one of them may stand in queue.
Jockeying	: If there are number of queues then one may leave one queue to join another.
Queue length	: No. of customers waiting in the queue.
Queuing system	: System consisting arrival of customers, waiting in queue, picked up for service according to a certain discipline, being serviced and departure of customers.
Service station	: Point where service is provided
Customer	: Person or unit arriving at a station for service. Customer may be a machine or person.
Waiting time	: Time a customer spends in the queue before being serviced.

9.11 QUESTIONS FOR DISCUSSION

1. **Write True or False against each statement:**

- (a) Parallel channel means, a number of channels providing identical services.
- (b) Queuing theory is concerned with the decision making process.
- (c) Customer decide not to joins the queue is Reneging.
- (d) The line that form in front of service facilities is called a queue.
- (e) Random arrival mean when there is how service point.

2. **Briefly comment on the following statements:**

- (a) In designing a queuing system, the system should balance service to customers.
- (b) Queuing theory deals with problem that involve waiting.
- (c) Most of queuing models are quite complex and cannot be easily understood.
- (d) In a single channel facility the output of the queue does not pose any problem.
- (e) The object of the queuing theory is to achieve a good economic balance and also to minimise the total waiting & service cost.

3. **Fill in the blank:**

- (a) The arrival time for any customer is _____
- (b) The most common queue discipline _____
- (c) If there is queue then _____ has to wait for same lines.
- (d) Cost associated with service or the facility are known as _____
- (e) Waiting time is a time which customer spends in the _____

4. **Write Short Notes:**

- (a) Queue discipline
- (b) Service mechanism is queuing theory
- (c) Queuing system

9.12 TERMINAL QUESTIONS

1. What is the queuing theory?
2. Define arrival rate and service rate.
3. Explain the characteristics of MM1 queuing model.
4. Briefly explain Service Mechanism and Queue Discipline.
5. What is system of steady-state?

Exercise Problems

1. A Bank operates a single facility ATM machine. Customers arrive at the rate of 10 customers per hour according to Poisson probability distribution. The time taken for an ATM transaction is exponential which means 3 minutes on an average. Find the following:
 - (a) Average waiting time of a customer before service.
 - (b) Average number of customers in the system.
 - (c) Probability that the ATM is idle.
2. At an average 12 cars per hour arrive at a single-server, drive-in teller. The average service time for each customer is 4 minutes, and the arrivals and services are Poisson and exponentially distributed respectively. Answer the following questions:
 - (a) What is the proportion that the teller is idle?
 - (b) What is the time spent by a customer to complete his transaction?
 - (c) What is the probability that an arriving car need not wait to take-up service?
3. At a single facility security check at an airport, passengers arrive at the checkpoint on an average of 8 passengers per minute and follows a Poisson probability distribution. The checking time for a customer entering security check area takes 10 passengers per minute and follows an exponential probability distribution. Determine the following:
 - (a) On an average, how many passengers are waiting in queue to enter the checkpoint?
 - (b) On an average, what is the time taken by a customer leaving the checkpoint?
4. In a college computer lab, computers are interconnected to one laser printer. The printer receives data files for printing from these 25 computers interconnected to it. The printer prints the files received from these 25 computers at the rate of 5 data files per minute. The average time required to print a data file is 6 minutes. Assuming the arrivals are Poisson distributed and service times are exponentially distributed, determine
 - (a) What is the probability that the printer is busy?
 - (b) On an average, how much time must a computer operator wait to take a print-out?
 - (c) On an average, what is the expected number of operators that will be waiting to take a print-out?
5. Skyline pizza is a famous restaurant operating a number of outlets. The restaurant uses a toll-free telephone number to book pizzas at any of its outlets. It was found that an average of 15 calls are received per hour and the average time to handle each call is 2.5 minutes. Determine the following:

- (a) What is the average waiting time of an incoming caller?
 - (b) What is the probability that a caller gets connected immediately?
 - (c) If the restaurant manager feels that average waiting time of a caller is more than 5 minutes, will lead to customer loss and the restaurant will have to go in for a second toll free facility, what should be the new arrival rate in order to justify another facility?
6. From historical data, a two-wheeler service station observe that bikes arrive only for water wash is at the rate of 7 per hour per 8 hour shift. The manager has a record that it takes 5 minutes for water service and another 2 minutes for greasing and general check. Assuming that one bike is washed at a time, find the following:
- (a) Average number of bikes in line.
 - (b) Average time a bike waits before it is washed.
 - (c) Average time a bike spends in the system.
 - (d) Utilization rate of the bike wash.
 - (e) Probability that no bikes are in the system.
7. In a department at store, an automated coffee vending machine is installed. Customers arrive at a rate of 3 per minute and it takes average time of 10 seconds to dispense a cup of coffee:
- (a) Determine the number of customers in the queue.
 - (b) Determine the waiting time of a customer.
 - (c) Find the probability that there are exactly 10 customers in the system.
8. In a toll gate, vehicles arrive at a rate of 120 per hour. An average time for a vehicle to get a pass is 25 seconds. The arrivals follow a Poisson distribution and service times follow an exponential distribution. (a) Find the average number of vehicles waiting and the idle time of the check-post. (b) If the idle time of the check post is less than 10%, the check-post authorities will install a second gate. Suggest whether a second gate is necessary ?
9. A hospital has an X-ray lab where patients (both in-patient and out-patient) arrive at a rate of 5 per minute. Due to variation in requirement, the time taken for one patient is 3 minutes and follows an exponential distribution. (a) What is the probability that the system is busy? and (b) What is the probability that nobody is in the system?
10. In the production shop of a company breakdown of the machine is found to be Poisson with an average rate of 3 machines per hour. Breakdown time at one machine costs Rs. 40 per hour to the company. There are two choices before the company for hiring the repairmen, one of the repairmen is slow but cheap, the other is fast but expensive. The slow-cheap repairman demands Rs. 20 per hour and will repair the breakdown machine exponentially at the rate of 4 per hour. The fast expensive repairman demands Rs. 30 per hour and will repair exponentially on an average rate of Rs.6 per hour. Which repairman should be hired?

9.13 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

- | | | |
|----------------------|-------------------------------------|-----------|
| 1. (a) True | (b) True | (c) False |
| (d) True | (e) False | |
| 3. (a) unpredictable | (b) [First come first serve (FCFS)] | |
| (c) customer | (d) service cost | (e) queue |

9.14 SUGGESTED READINGS

- T.L. Satty, *Elements of Queueing*, New York McGraw Hill Theory.
- A.M. Lee, *Applied Queueing thoery*.
- Cooper, R.B., *Introduction to Queueing theory*, New York MacMillan Co.
- Morse, Philip M.Ques, *Inventories & maintenance*, New York John Wiley & Sons.
- Panieo, J.A., *Queueing Theory: A Study of Waiting Line of Business, Economics, Science*, Englewood, No. 3, Prentice-Hall.
- Bhat, U.N. “*The Value of Queueing Theory – A Rejoinder Interface*, Vol. 8, No. 3 pp. 27-78.
- Byrd. J., *The value of queueing & Interfaces*, Vol. 8, No. 3, pp. 22-26.
- Render, B & Stain R.M., *Cases & readings in quantitative analysis*, Boston: Allyn & Bacon, 1982.
- Graff G, “*Simple Queueing thoery saves Unneccessary Equipment*”, Industrial Engineering, Vol. 3.
- Paul R.I., Stevens R.E., *Staffing service activities with waiting the models decision*, Science, Vol. 2

Unit-IV

LESSON

10

PROBABILITY

CONTENTS

- 10.0 Aims and Objectives
- 10.1 Introduction
- 10.2 Classical Definition of Probability
- 10.3 Counting Techniques
- 10.4 Statistical or Empirical Definition of Probability
- 10.5 Axiomatic or Modern Approach to Probability
- 10.6 Theorems on Probability-I
- 10.7 Theorems on Probability-II
- 10.8 Let us Sum Up
- 10.9 Lesson-end Activity
- 10.10 Keywords
- 10.11 Questions for Discussion
- 10.12 Terminal Questions
- 10.13 Model Answers to Questions for Discussion
- 10.14 Suggested Readings

10.0 AIMS AND OBJECTIVES

The probability, theoretical probability distribution and probability distribution of random variable in the three important interrelated trades which we are going to discuss in this head (Unit IV). As we know that probability associated with the occurrence of various events are determined by specifying the condition of a random experiments.

10.1 INTRODUCTION

The concept of probability originated from the analysis of the games of chance in the 17th century. Now the subject has been developed to the extent that it is very difficult to imagine a discipline, be it from social or natural sciences, that can do without it. The theory of probability is a study of *Statistical or Random Experiments*. It is the backbone of *Statistical Inference* and *Decision Theory* that are essential tools of the analysis of most of the modern business and economic problems.

Often, in our day-to-day life, we hear sentences like 'it may rain today', 'Mr X has fifty-fifty chances of passing the examination', 'India may win the forthcoming cricket match against Sri Lanka', 'the chances of making profits by investing in shares of company A are very bright', etc. Each of the above sentences involves an element of uncertainty.

A phenomenon or an experiment which can result into more than one possible outcome, is called a random phenomenon or random experiment or statistical experiment. Although, we may be aware of all the possible outcomes of a random experiment, it is not possible to predetermine the outcome associated with a particular experimentation or trial.

Consider, for example, the toss of a coin. The result of a toss can be a head or a tail, therefore, it is a random experiment. Here we know that either a head or a tail would occur as a result of the toss, however, it is not possible to predetermine the outcome. With the use of probability theory, it is possible to assign a quantitative measure, to express the extent of uncertainty, associated with the occurrence of each possible outcome of a random experiment.

10.2 CLASSICAL DEFINITION OF PROBABILITY

This definition, also known as the mathematical definition of probability, was given by J. Bernoulli. With the use of this definition, the probabilities associated with the occurrence of various events are determined by specifying the conditions of a random experiment. It is because of this that the classical definition is also known as 'a priori' definition of probability.

Definition

If n is the number of equally likely, mutually exclusive and exhaustive outcomes of a random experiment out of which m outcomes are favourable to the occurrence of an event A, then the probability that A occurs, denoted by $P(A)$, is given by :

$$P(A) = \frac{\text{Number of outcomes favourable to } A}{\text{Number of exhaustive outcomes}} = \frac{m}{n}$$

Various terms used in the above definition are explained below :

1. ***Equally likely outcomes:*** The outcomes of random experiment are said to be equally likely or equally probable if the occurrence of none of them is expected in preference to others. For example, if an unbiased coin is tossed, the two possible outcomes, a head or a tail are equally likely.
2. ***Mutually exclusive outcomes:*** Two or more outcomes of an experiment are said to be mutually exclusive if the occurrence of one of them precludes the occurrence of all others in the same trial. For example, the two possible outcomes of toss of a coin are mutually exclusive. Similarly, the occurrences of the numbers 1, 2, 3, 4, 5, 6 in the roll of a six faced die are mutually exclusive.
3. ***Exhaustive outcomes:*** It is the totality of all possible outcomes of a random experiment. The number of exhaustive outcomes in the roll of a die are six. Similarly, there are 52 exhaustive outcomes in the experiment of drawing a card from a pack of 52 cards.
4. ***Event:*** The occurrence or non-occurrence of a phenomenon is called an event. For example, in the toss of two coins, there are four exhaustive outcomes, viz. (H, H), (H, T), (T, H), (T, T). The events associated with this experiment can be defined in a number of ways. For example, (i) the event of occurrence of head on both the coins, (ii) the event of occurrence of head on at least one of the two coins, (iii) the event of non-occurrence of head on the two coins, etc.

An event can be simple or composite depending upon whether it corresponds to a single outcome of the experiment or not. In the example, given above, the event defined by (i) is simple, while those defined by (ii) and (iii) are composite events.

Example 1: What is the probability of obtaining a head in the toss of an unbiased coin?

Solution: This experiment has two possible outcomes, i.e., occurrence of a head or tail. These two outcomes are mutually exclusive and exhaustive. Since the coin is given to be unbiased, the two outcomes are equally likely. Thus, all the conditions of the classical definition are satisfied.

No. of cases favourable to the occurrence of head = 1

Probability

No. of exhaustive cases = 2

$$\therefore \text{Probability of obtaining head } P(H) = \frac{1}{2}.$$

Example 2: What is the probability of obtaining at least one head in the simultaneous toss of two unbiased coins?

Solution: The equally likely, mutually exclusive and exhaustive outcomes of the experiment are (H, H), (H, T), (T, H) and (T, T), where H denotes a head and T denotes a tail. Thus, $n = 4$.

Let A be the event that at least one head occurs. This event corresponds to the first three outcomes of the random experiment. Therefore, $m = 3$.

Hence, probability that A occurs, i.e., $P(A) = \frac{3}{4}$.

Example 3: Find the probability of obtaining an odd number in the roll of an unbiased die.

Solution: The number of equally likely, mutually exclusive and exhaustive outcomes, i.e., $n = 6$. There are three odd numbers out of the numbers 1, 2, 3, 4, 5 and 6. Therefore, $m = 3$.

Thus, probability of occurrence of an odd number = $\frac{3}{6} = \frac{1}{2}$.

Example 4: What is the chance of drawing a face card in a draw from a pack of 52 well-shuffled cards?

Solution: Total possible outcomes $n = 52$.

Since the pack is well-shuffled, these outcomes are equally likely. Further, since only one card is to be drawn, the outcomes are mutually exclusive.

There are 12 face cards, $\therefore m = 12$.

Thus, probability of drawing a face card = $\frac{12}{52} = \frac{3}{13}$.

Example 5: What is the probability that a leap year selected at random will contain 53 Sundays?

Solution: A leap year has 366 days. It contains 52 complete weeks, i.e., 52 Sundays. The remaining two days of the year could be anyone of the following pairs :

(Monday, Tuesday), (Tuesday, Wednesday), (Wednesday, Thursday), (Thursday, Friday), (Friday, Saturday), (Saturday, Sunday), (Sunday, Monday). Thus, there are seven possibilities out of which last two are favourable to the occurrence of 53rd Sunday.

Hence, the required probability = $\frac{2}{7}$.

Example 6: Find the probability of throwing a total of six in a single throw with two unbiased dice.

Solution: The number of exhaustive cases $n = 36$, because with two dice all the possible outcomes are :

- (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),
- (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),
- (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),
- (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),
- (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),
- (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6).

Out of these outcomes the number of cases favourable to the event A of getting 6 are : (1, 5), (2, 4), (3, 3), (4, 2), (5, 1). Thus, we have $m = 5$.

$$\therefore P(A) = \frac{5}{36}$$

Example 7: A bag contains 15 tickets marked with numbers 1 to 15. One ticket is drawn at random. Find the probability that:

- (i) the number on it is greater than 10,
- (ii) the number on it is even,
- (iii) the number on it is a multiple of 2 or 5.

Solution: Number of exhaustive cases $n = 15$

- (i) Tickets with number greater than 10 are 11, 12, 13, 14 and 15. Therefore, $m = 5$ and

$$\text{hence the required probability} = \frac{5}{15} = \frac{1}{3}$$

- (ii) Number of even numbered tickets $m = 7$

$$\therefore \text{Required probability} = \frac{7}{15}$$

- (iii) The multiple of 2 are : 2, 4, 6, 8, 10, 12, 14 and the multiple of 5 are : 5, 10, 15.
 $\therefore m = 9$ (note that 10 is repeated in both multiples will be counted only once).

$$\text{Thus, the required probability} = \frac{9}{15} = \frac{3}{5}$$

10.3 COUNTING TECHNIQUES

Counting techniques or combinatorial methods are often helpful in the enumeration of total number of outcomes of a random experiment and the number of cases favourable to the occurrence of an event.

Fundamental Principle of Counting

If the first operation can be performed in any one of the m ways and then a second operation can be performed in any one of the n ways, then both can be performed together in any one of the $m \times n$ ways.

This rule can be generalised. If first operation can be performed in any one of the n_1 ways, second operation in any one of the n_2 ways, k th operation in any one of the n_k ways, then together these can be performed in any one of the $n_1 \times n_2 \times \dots \times n_k$ ways.

Permutation

A permutation is an arrangement of a given number of objects in a definite order.

- (a) **Permutations of n objects:** The total number of permutations of n distinct objects is $n!$. Using symbols, we can write ${}^n P_n = n!$, (where n denotes the permutations of n objects, all taken together).

Let us assume there are n persons to be seated on n chairs. The first chair can be occupied by any one of the n persons and hence, there are n ways in which it can be occupied. Similarly, the second chair can be occupied in $n - 1$ ways and so on. Using the fundamental principle of counting, the total number of ways in which n chairs can be occupied by n persons or the permutations of n objects taking all at a time is given by :

$${}^n P_n = n(n - 1)(n - 2) \dots 3.2.1 = n!$$

- (b) **Permutations of n objects taking r at a time:** In terms of the example, considered above, now we have n persons to be seated on r chairs, where $r \leq n$.

Thus, ${}^n P_r = n(n - 1)(n - 2) \dots [n - (r - 1)] = n(n - 1)(n - 2) \dots (n - r + 1)$.

$${}^n P_r = \frac{n(n-1)(n-2) \dots (n-r+1)(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

- (c) **Permutations of n objects taking r at a time when any object may be repeated any number of times:** Here, each of the r places can be filled in n ways. Therefore, total number of permutations is n^r .
- (d) **Permutations of n objects in a circular order:** Suppose that there are three persons A, B and C, to be seated on the three chairs 1, 2 and 3, in a circular order. Then, the following three arrangements are identical:

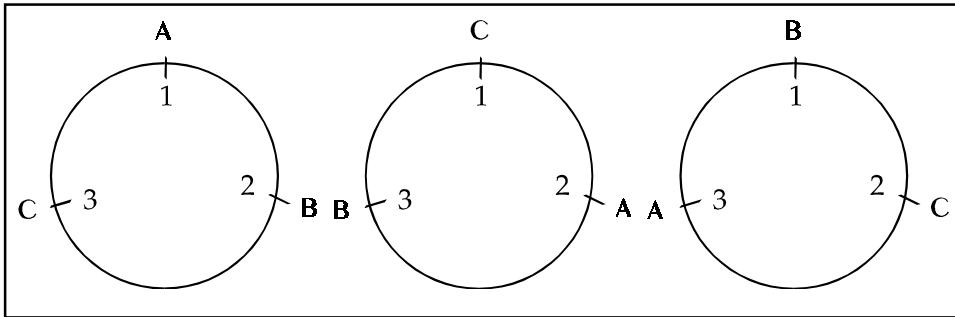


Figure 10.1

Similarly, if n objects are seated in a circle, there will be n identical arrangements of the above type. Thus, in order to obtain distinct permutation of n objects in circular order we divide ${}^n P_n$ by n , where ${}^n P_n$ denotes number of permutations in a row.

Hence, the number of permutations in a circular order $\frac{n!}{n} = (n-1)!$

- (e) **Permutations with restrictions:** If out of n objects n_1 are alike of one kind, n_2 are alike of another kind, n_k are alike, the number of permutations are $\frac{n!}{n_1! n_2! \dots n_k!}$. Since permutation of n_i objects, which are alike, is only one ($i = 1, 2, \dots, k$). Therefore, $n!$ is to be divided by $n_1!, n_2!, \dots, n_k!$, to get the required permutations.

Example 8: What is the total number of ways of simultaneous throwing of (i) 3 coins, (ii) 2 dice and (iii) 2 coins and a die?

Solution:

- (i) Each coin can be thrown in any one of the two ways, i.e., a head or a tail, therefore, the number of ways of simultaneous throwing of 3 coins $= 2^3 = 8$.
- (ii) Similarly, the total number of ways of simultaneous throwing of two dice is equal to $6^2 = 36$ and
- (iii) The total number of ways of simultaneous throwing of 2 coins and a die is equal to $2^2 \times 6 = 24$.

Example 9: A person can go from Delhi to Port-Blair via Allahabad and Calcutta using following mode of transport :

Delhi to Allahabad	Allahabad to Calcutta	Calcutta to Port-Blair
By Rail	By Rail	By Air
By Bus	By Bus	By Ship
By Car	By Car	
By Air	By Air	

In how many different ways the journey can be planned?

Solution: The journey from Delhi to Port-Blair can be treated as three operations; From Delhi to Allahabad, from Allahabad to Calcutta and from Calcutta to Port-Blair. Using the fundamental principle of counting, the journey can be planned in $4 \times 4 \times 2 = 32$ ways.

Example 10: In how many ways the first, second and third prize can be given to 10 competitors?

Solution: There are 10 ways of giving first prize, nine ways of giving second prize and eight ways of giving third prize. Therefore, total no. ways is $10 \times 9 \times 8 = 720$.

Alternative method:

$$\text{Here } n = 10 \text{ and } r = 3, \therefore {}^{10}P_3 = \frac{10!}{(10-3)!} = 720$$

Example 11:

- (a) There are 5 doors in a room. In how many ways can three persons enter the room using different doors?
- (b) A lady is asked to rank 5 types of washing powders according to her preference. Calculate the total number of possible rankings.
- (c) In how many ways 6 passengers can be seated on 15 available seats.
- (d) If there are six different trains available for journey between Delhi to Kanpur, calculate the number of ways in which a person can complete his return journey by using a different train in each direction.
- (e) In how many ways President, Vice-President, Secretary and Treasurer of an association can be nominated at random out of 130 members?

Solution:

- (a) The first person can use any of the 5 doors and hence can enter the room in 5 ways. Similarly, the second person can enter in 4 ways and third person can enter in 3 ways. Thus, the total number of ways is ${}^5P_3 = \frac{5!}{2!} = 60$.
- (b) Total number of rankings are ${}^5P_5 = \frac{5!}{0!} = 120$. (Note that $0! = 1$)
- (c) Total number of ways of seating 6 passengers on 15 seats are ${}^{15}P_6 = \frac{15!}{9!} = 36,03,600$.
- (d) Total number of ways of performing return journey, using different train in each direction are $6 \times 5 = 30$, which is also equal to 6P_2 .
- (e) Total number of ways of nominating for the 4 post of association are ${}^{130}P_4 = \frac{130!}{126!} = 27,26,13,120$.

Example 12: Three prizes are awarded each for getting more than 80% marks, 98% attendance and good behaviour in the college. In how many ways the prizes can be awarded if 15 students of the college are eligible for the three prizes?

Solution: Note that all the three prizes can be awarded to the same student. The prize for getting more than 80% marks can be awarded in 15 ways, prize for 90% attendance can be awarded in 15 ways and prize for good behaviour can also be awarded in 15 ways.

Thus, the total number of ways is $n^r = 15^3 = 3,375$.

Example 13:

- (a) In how many ways can the letters of the word EDUCATION be arranged?
- (b) In how many ways can the letters of the word STATISTICS be arranged?

- (c) In how many ways can 20 students be allotted to 4 tutorial groups of 4, 5, 5 and 6 students respectively?
- (d) In how many ways 10 members of a committee can be seated at a round table if (i) they can sit anywhere (ii) president and secretary must not sit next to each other?

Probability

Solution:

- (a) The given word EDUCATION has 9 letters. Therefore, number of permutations of 9 letters is $9! = 3,62,880$.
- (b) The word STATISTICS has 10 letters in which there are 3S's, 3T's, 2I's, 1A and 1C. Thus, the required number of permutations $\frac{10!}{3!3!2!1!1!} = 50,400$.
- (c) Required number of permutations $\frac{20!}{4!5!5!6!} = 9,77,72,87,522$
- (d) (i) Number of permutations when they can sit anywhere $= (10-1)! = 9! = 3,62,880$.
(ii) We first find the number of permutations when president and secretary must sit together. For this we consider president and secretary as one person. Thus, the number of permutations of 9 persons at round table $= 8! = 40,320$.
∴ The number of permutations when president and secretary must not sit together $= 3,62,880 - 40,320 = 3,22,560$.

Example 14:

- (a) In how many ways 4 men and 3 women can be seated in a row such that women occupy the even places?
- (b) In how many ways 4 men and 4 women can be seated such that men and women occupy alternative places?

Solution:

- (a) 4 men can be seated in $4!$ ways and 3 women can be seated in $3!$ ways. Since each arrangement of men is associated with each arrangement of women, therefore, the required number of permutations $= 4! 3! = 144$.
- (b) There are two ways in which 4 men and 4 women can be seated

$$\text{MWMWMWMWMW or WMWMWMWMWM}$$

$$\therefore \text{The required number of permutations} = 2 \cdot 4! 4! = 1,152$$

Example 15: There are 3 different books of economics, 4 different books of commerce and 5 different books of statistics. In how many ways these can be arranged on a shelf when

- (a) all the books are arranged at random,
(b) books of each subject are arranged together,
(c) books of only statistics are arranged together, and
(d) books of statistics and books of other subjects are arranged together?

Solution:

- (a) The required number of permutations $= 12!$
- (b) The economics books can be arranged in $3!$ ways, commerce books in $4!$ ways and statistics book in $5!$ ways. Further, the three groups can be arranged in $3!$ ways. ∴ The required number of permutations $= 3! 4! 5! 3! = 1,03,680$.
- (c) Consider 5 books of statistics as one book. Then 8 books can be arranged in $8!$ ways and 5 books of statistics can be arranged among themselves in $5!$ ways.
∴ The required number of permutations $= 8! 5! = 48,38,400$.

- (d) There are two groups which can be arranged in $2!$ ways. The books of other subjects can be arranged in $7!$ ways and books of statistics can be arranged in $5!$ ways. Thus, the required number of ways = $2! \cdot 7! \cdot 5! = 12,09,600$.

Combination

When no attention is given to the order of arrangement of the selected objects, we get a combination. We know that the number of permutations of n objects taking r at a time is ${}^n P_r$. Since r objects can be arranged in $r!$ ways, therefore, there are $r!$ permutations corresponding to one combination. Thus, the number of combinations of n objects taking r at a time, denoted by ${}^n C_r$, can be obtained by dividing ${}^n P_r$ by $r!$, i.e.,

$${}^n C_r = \frac{{}^n P_r}{r!} = \frac{n!}{r!(n-r)!}.$$

- Note:**
- (a) Since ${}^n C_r = {}^n C_{n-r}$, therefore, ${}^n C_r$ is also equal to the combinations of n objects taking $(n - r)$ at a time.
 - (b) The total number of combinations of n distinct objects taking 1, 2, ..., n respectively, at a time is ${}^n C_1 + {}^n C_2 + \dots + {}^n C_n = 2^n - 1$.

Example 16:

- (a) In how many ways two balls can be selected from 8 balls?
- (b) In how many ways a group of 12 persons can be divided into two groups of 7 and 5 persons respectively?
- (c) A committee of 8 teachers is to be formed out of 6 science, 8 arts teachers and a physical instructor. In how many ways the committee can be formed if
 1. Any teacher can be included in the committee.
 2. There should be 3 science and 4 arts teachers on the committee such that
 - (i) any science teacher and any arts teacher can be included,
 - (ii) one particular science teacher must be on the committee,
 - (iii) three particular arts teachers must not be on the committee?

Solution:

- (a) 2 balls can be selected from 8 balls in ${}^8 C_2 = \frac{8!}{2!6!} = 28$ ways.
- (b) Since ${}^n C_r = {}^n C_{n-r}$, therefore, the number of groups of 7 persons out of 12 is also equal to the number of groups of 5 persons out of 12. Hence, the required number of groups is ${}^{12} C_7 = \frac{12!}{7!5!} = 792$.

Alternative Method: We may regard 7 persons of one type and remaining 5 persons of another type. The required number of groups are equal to the number of permutations of 12 persons where 7 are alike of one type and 5 are alike of another type.

- (c) 1. 8 teachers can be selected out of 15 in ${}^{15} C_8 = \frac{15!}{8!7!} = 6,435$ ways.
- 2. (i) 3 science teachers can be selected out of 6 teachers in ${}^6 C_3$ ways and 4 arts teachers can be selected out of 8 in ${}^8 C_4$ ways and the physical instructor can be selected in ${}^1 C_1$ way. Therefore, the required number of ways = ${}^6 C_3 \times {}^8 C_4 \times {}^1 C_1 = 20 \times 70 \times 1 = 1400$.

- (ii) 2 additional science teachers can be selected in 5C_2 ways. The number of selections of other teachers is same as in (i) above. Thus, the required number of ways = ${}^5C_2 \times {}^8C_4 \times {}^1C_1 = 10 \times 70 \times 1 = 700$.
- (iii) 3 science teachers can be selected in 6C_3 ways and 4 arts teachers out of remaining 5 arts teachers can be selected in 5C_4 ways.

\therefore The required number of ways = ${}^6C_3 \times {}^5C_4 = 20 \times 5 = 100$.

Ordered Partitions

1. Ordered Partitions (distinguishable objects)

- (a) The total number of ways of putting n distinct objects into r compartments which are marked as 1, 2, r is equal to r^n .

Since first object can be put in any of the r compartments in r ways, second can be put in any of the r compartments in r ways and so on.

- (b) The number of ways in which n objects can be put into r compartments such that the first compartment contains n_1 objects, second contains n_2 objects and so on the rth compartment contains n_r objects, where $n_1 + n_2 + \dots + n_r = n$,

is given by $\frac{n!}{n_1! n_2! \dots n_r!}$.

To illustrate this, let $r = 3$. Then n_1 objects in the first compartment can be put in ${}^nC_{n_1}$ ways. Out of the remaining $n - n_1$ objects, n_2 objects can be put in the second compartment in ${}^{n-n_1}C_{n_2}$ ways. Finally the remaining $n - n_1 - n_2 = n_3$ objects can be put in the third compartment in one way. Thus, the required

number of ways is ${}^nC_{n_1} \times {}^{n-n_1}C_{n_2} = \frac{n!}{n_1! n_2! n_3!}$

2. Ordered Partitions (identical objects)

- (a) The total number of ways of putting n identical objects into r compartments marked as 1, 2, r, is ${}^n C_{r-1}$, where each compartment may have none or any number of objects.

We can think of n objects being placed in a row and partitioned by the $(r - 1)$ vertical lines into r compartments. This is equivalent to permutations of $(n + r - 1)$ objects out of which n are of one type and $(r - 1)$ of another type.

The required number of permutations are $\frac{(n+r-1)!}{n!(r-1)!}$, which is equal to

${}^{(n+r-1)}C_n$ or ${}^{(n+r-1)}C_{(r-1)}$.

- (b) The total number of ways of putting n identical objects into r compartments is ${}^{(n-r)+(r-1)}C_{(r-1)}$ or ${}^{(n-1)}C_{(r-1)}$, where each compartment must have at least one object.

In order that each compartment must have at least one object, we first put one object in each of the r compartments. Then the remaining $(n - r)$ objects can be placed as in (a) above.

- (c) The formula, given in (b) above, can be generalised. If each compartment is supposed to have at least k objects, the total number of ways is ${}^{(n-kr)+(r-1)}C_{(r-1)}$,

where $k = 0, 1, 2, \dots$ etc. such that $k < \frac{n}{r}$.

Example 17: 4 couples occupy eight seats in a row at random. What is the probability that all the ladies are sitting next to each other?

Solution: Eight persons can be seated in a row in $8!$ ways.

We can treat 4 ladies as one person. Then, five persons can be seated in a row in $5!$ ways. Further, 4 ladies can be seated among themselves in $4!$ ways.

$$\therefore \text{The required probability} = \frac{5!4!}{8!} = \frac{1}{14}$$

Example 18: 12 persons are seated at random (i) in a row, (ii) in a ring. Find the probabilities that three particular persons are sitting together.

Solution:

$$(i) \quad \text{The required probability} = \frac{10!3!}{12!} = \frac{1}{22}$$

$$(ii) \quad \text{The required probability} = \frac{9!3!}{11!} = \frac{3}{55}.$$

Example 19: 5 red and 2 black balls, each of different sizes, are randomly laid down in a row. Find the probability that

- (i) the two end balls are black,
- (ii) there are three red balls between two black balls and
- (iii) the two black balls are placed side by side.

Solution: The seven balls can be placed in a row in $7!$ ways.

- (i) The black can be placed at the ends in $2!$ ways and, in-between them, 5 red balls can be placed in $5!$ ways.

$$\therefore \text{The required probability} = \frac{2!5!}{7!} = \frac{1}{21}.$$

- (ii) We can treat BRRRB as one ball. Therefore, this ball along with the remaining two balls can be arranged in $3!$ ways. The sequence BRRRB can be arranged in $2! 3!$ ways and the three red balls of the sequence can be obtained from 5 balls in 5C_3 ways.

$$\therefore \text{The required probability} = \frac{3!2!3!}{7!} \times {}^5C_3 = \frac{1}{7}.$$

- (iii) The 2 black balls can be treated as one and, therefore, this ball along with 5 red balls can be arranged in $6!$ ways. Further, 2 black ball can be arranged in $2!$ ways.

$$\therefore \text{The required probability} = \frac{6!2!}{7!} = \frac{2}{7}$$

Example 20: Each of the two players, A and B, get 26 cards at random. Find the probability that each player has an equal number of red and black cards.

Solution: Each player can get 26 cards at random in ${}^{52}C_{26}$ ways.

In order that a player gets an equal number of red and black cards, he should have 13 cards of each colour, note that there are 26 red cards and 26 black cards in a pack of playing cards. This can be done in ${}^{26}C_{13} \times {}^{26}C_{13}$ ways. Hence, the required

$$\text{probability} = \frac{{}^{26}C_{13} \times {}^{26}C_{13}}{{}^{52}C_{26}}.$$

Example 21: 8 distinguishable marbles are distributed at random into 3 boxes marked as 1, 2 and 3. Find the probability that they contain 3, 4 and 1 marbles respectively.

Solution: Since the first, second 8th marble, each, can go to any of the three boxes in 3 ways, the total number of ways of putting 8 distinguishable marbles into three boxes is 3^8 .

The number of ways of putting the marbles, so that the first box contains 3 marbles, second contains 4 and the third contains 1, are $\frac{8!}{3!4!1!}$

$$\therefore \text{The required probability} = \frac{8!}{3!4!1!} \times \frac{1}{3^8} = \frac{280}{6561}.$$

Example 22: 12 'one rupee' coins are distributed at random among 5 beggars A, B, C, D and E. Find the probability that :

- (i) They get 4, 2, 0, 5 and 1 coins respectively.
- (ii) Each beggar gets at least two coins.
- (iii) None of them goes empty handed.

Solution: The total number of ways of distributing 12 one rupee coins among 5 beggars are ${}^{12+5-1}C_{5-1} = {}^{16}C_4 = 1820$.

- (i) Since the distribution 4, 2, 0, 5, 1 is one way out of 1820 ways, the required probability

$$= \frac{1}{1820}.$$

- (ii) After distributing two coins to each of the five beggars, we are left with two coins, which can be distributed among five beggars in ${}^{2+5-1}C_{5-1} = {}^6C_4 = 15$ ways.

$$\therefore \text{The required probability} = \frac{15}{1820} = \frac{3}{364}$$

- (iii) No beggar goes empty handed if each gets at least one coin. 7 coins, that are left after giving one coin to each of the five beggars, can be distributed among five beggars in ${}^{7+5-1}C_{5-1} = {}^{11}C_4 = 330$ ways.

$$\therefore \text{The required probability} = \frac{330}{1820} = \frac{33}{182}$$

10.4 STATISTICAL OR EMPIRICAL DEFINITION OF PROBABILITY

The scope of the classical definition was found to be very limited as it failed to determine the probabilities of certain events in the following circumstances :

- (i) When n, the exhaustive outcomes of a random experiment is infinite.
- (ii) When actual value of n is not known.
- (iii) When various outcomes of a random experiment are not equally likely.
- (iv) This definition doesn't lead to any mathematical treatment of probability.

In view of the above shortcomings of the classical definition, an attempt was made to establish a correspondence between relative frequency and the probability of an event when the total number of trials become sufficiently large.

Definition (R. Von Mises)

If an experiment is repeated n times, under essentially the identical conditions and, if, out of these trials, an event A occurs m times, then the probability that

A occurs is given by $P(A) = \lim_{n \rightarrow \infty} \frac{m}{n}$, provided the limit exists.

This definition of probability is also termed as the empirical definition because the probability of an event is obtained by actual experimentation.

Although, it is seldom possible to obtain the limit of the relative frequency, the ratio $\frac{m}{n}$ can be regarded as a good approximation of the probability of an event for large values of n .

This definition also suffers from the following shortcomings :

- (i) The conditions of the experiment may not remain identical, particularly when the number of trials is sufficiently large.
- (ii) The relative frequency, $\frac{m}{n}$, may not attain a unique value no matter how large is the total number of trials.
- (iii) It may not be possible to repeat an experiment a large number of times.
- (iv) Like the classical definition, this definition doesn't lead to any mathematical treatment of probability.

10.5 AXIOMATIC OR MODERN APPROACH TO PROBABILITY

This approach was introduced by the Russian mathematician, A. Kolmogorov in 1930s. In his book, 'Foundations of Probability' published in 1933, he introduced probability as a function of the outcomes of an experiment, under certain restrictions. These restrictions are known as Postulates or Axioms of probability theory. Before discussing the above approach to probability, we shall explain certain concepts that are necessary for its understanding.

Sample Space

It is the set of all possible outcomes of a random experiment. Each element of the set is called a *sample point* or a *simple event* or an *elementary event*. The sample space of a random experiment is denoted by S and its elements are denoted by e_i , where $i = 1, 2, \dots, n$. Thus, a sample space having n elements can be written as :

$$S = \{e_1, e_2, \dots, e_n\}.$$

If a random experiment consists of rolling a six faced die, the corresponding sample space consists of 6 elementary events. Thus, $S = \{1, 2, 3, 4, 5, 6\}$.

Similarly, in the toss of a coin $S = \{H, T\}$.

The elements of S can either be single elements or ordered pairs. For example, if two coins are tossed, each element of the sample space would consist of the set of ordered pairs, as shown below :

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

Finite and Infinite Sample Space

A sample space consisting of finite number of elements is called a *finite sample space*, while if the number of elements is infinite, it is called an *infinite sample space*. The sample spaces discussed so far are examples of finite sample spaces. As an example of infinite sample space, consider repeated toss of a coin till a head appears. Various elements of the sample space would be :

$$S = \{(H), (T, H), (T, T, H), \dots\}.$$

A discrete sample space consists of finite or countably infinite number of elements. The sample spaces, discussed so far, are some examples of discrete sample spaces. Contrary to this, a continuous sample space consists of an uncountable number of elements. This type of sample space is obtained when the result of an experiment is a measurement on continuous scale like measurements of weight, height, area, volume, time, etc.

Event

An event is any subset of a sample space. In the experiment of roll of a die, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$. It is possible to define various events on this sample space, as shown below :

Let A be the event that an odd number appears on the die. Then $A = \{1, 3, 5\}$ is a subset of S. Further, let B be the event of getting a number greater than 4. Then $B = \{5, 6\}$ is another subset of S. Similarly, if C denotes an event of getting a number 3 on the die, then $C = \{3\}$.

It should be noted here that the events A and B are composite while C is a *simple or elementary event*.

Occurrence of an Event

An event is said to have occurred whenever the outcome of the experiment is an element of its set. For example, if we throw a die and obtain 5, then both the events A and B, defined above, are said to have occurred.

It should be noted here that the sample space is certain to occur since the outcome of the experiment must always be one of its elements.

Definition of Probability (Modern Approach)

Let S be a sample space of an experiment and A be any event of this sample space. The probability of A, denoted by $P(A)$, is defined as a real value set function which associates a real value corresponding to a subset A of the sample space S. In order that $P(A)$ denotes a probability function, the following rules, popularly known as axioms or postulates of probability, must be satisfied.

Axiom I : For any event A in sample space S, we have $0 \leq P(A) \leq 1$.

Axiom II : $P(S) = 1$.

Axiom III : If A_1, A_2, \dots, A_k are k mutually exclusive events (i.e., $\bigcap_{i \neq j} A_i = \emptyset$, where \emptyset denotes a null set) of the sample space S, then

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = \sum_{i=1}^k P(A_i)$$

The first axiom implies that the probability of an event is a non-negative number less than or equal to unity. The second axiom implies that the probability of an event that is certain to occur must be equal to unity. Axiom III gives a basic rule of addition of probabilities when events are mutually exclusive.

The above axioms provide a set of basic rules that can be used to find the probability of any event of a sample space.

Probability of an Event

Let there be a sample space consisting of n elements, i.e., $S = \{e_1, e_2, \dots, e_n\}$. Since the elementary events e_1, e_2, \dots, e_n are mutually exclusive, we have, according to axiom

III, $P(S) = \sum_{i=1}^n P(e_i)$. Similarly, if $A = \{e_1, e_2, \dots, e_m\}$ is any subset of S consisting of m

elements, where $m \leq n$, then $P(A) = \sum_{i=1}^m P(e_i)$. Thus, *the probability of a sample space or an event is equal to the sum of probabilities of its elementary events*.

It is obvious from the above that the probability of an event can be determined if the probabilities of elementary events, belonging to it, are known.

The Assignment of Probabilities to various Elementary Events

The assignment of probabilities to various elementary events of a sample space can be done in any one of the following three ways :

1. **Using Classical Definition:** We know that various elementary events of a random experiment, under the classical definition, are equally likely and, therefore, can be assigned equal probabilities. Thus, if there are n elementary events in the sample

space of an experiment and in view of the fact that $P(S) = \sum_{i=1}^n P(e_i) = 1$ (from axiom II), we can assign a probability equal to $\frac{1}{n}$ to every elementary event or,

using symbols, we can write $P(e_i) = \frac{1}{n}$ for $i = 1, 2, \dots, n$.

Further, if there are m elementary events in an event A, we have,

$$P(A) = \frac{1}{n} \quad \frac{1}{n} \quad \dots \quad \frac{1}{n} \quad (\text{m times}) \quad \frac{m}{n} = \frac{n(A), \text{i.e., number of elements in } A}{n(S), \text{i.e., number of elements in } S}$$

We note that the above expression is similar to the formula obtained under classical definition.

2. **Using Statistical Definition:** Using this definition, the assignment of probabilities to various elementary events of a sample space can be done by repeating an experiment a large number of times or by using the past records.
3. **Subjective Assignment:** The assignment of probabilities on the basis of the statistical and the classical definitions is *objective*. Contrary to this, it is also possible to have *subjective assignment* of probabilities. Under the subjective assignment, the probabilities to various elementary events are assigned on the basis of the expectations or the degree of belief of the statistician. These probabilities, also known as *personal probabilities*, are very useful in the analysis of various business and economic problems.

It is obvious from the above that the *Modern Definition* of probability is a general one which includes the classical and the statistical definitions as its particular cases. Besides this, it provides a set of mathematical rules that are useful for further mathematical treatment of the subject of probability.

Check Your Progress 10.1

- 1 Explain Exhaustive outcomes with examples.
2. What are combinational methods?

- Notes:**
- (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-
-

10.6 THEOREMS ON PROBABILITY - I

Theorem 1: $P(\phi) = 0$, where ϕ is a null set.

Proof: For a sample space S of an experiment, we can write $S \cup \phi = S$.

Taking probability of both sides, we have $P(S \cup \phi) = P(S)$.

Since S and ϕ are mutually exclusive, using axiom III, we can write $P(S) + P(\phi) = P(S)$. Hence, $P(\phi) = 0$.

Theorem 2: $P(\bar{A}) = 1 - P(A)$, where \bar{A} is compliment of A .

Proof: Let A be any event in the sample space S . We can write

$$A \cup \bar{A} = S \text{ or } P(A \cup \bar{A}) = P(S)$$

Since A and \bar{A} are mutually exclusive, we can write

$$P(A) + P(\bar{A}) = P(S) = 1. \text{ Hence, } P(\bar{A}) = 1 - P(A).$$

Theorem 3: For any two events A and B in a sample space S

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

Proof: From the Venn diagram, we can write

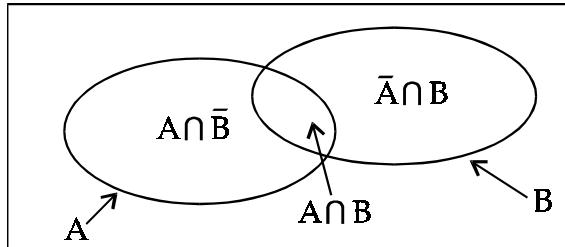
$$B = (\bar{A} \cap B) \cup (A \cap B) \text{ or}$$

$$P(B) = P[(\bar{A} \cap B) \cup (A \cap B)]$$

Since $(\bar{A} \cap B)$ and $(A \cap B)$ are mutually exclusive, we have

$$P(B) = P(\bar{A} \cap B) + P(A \cap B)$$

$$\text{or } P(\bar{A} \cap B) = P(B) - P(A \cap B).$$



Venn Diagram

Figure. 10.2

Similarly, it can be shown that

$$P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

Theorem 4: (Addition of Probabilities):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Proof: From the Venn diagram, given above, we can write

$$A \cup B = A \cup (\bar{A} \cap B) \text{ or } P(A \cup B) = P[A \cup (\bar{A} \cap B)]$$

Since A and $(\bar{A} \cap B)$ are mutually exclusive, we can write

$$P(A \cup B) = P(A) + P(\bar{A} \cap B)$$

Substituting the value of $P(\bar{A} \cap B)$ from theorem 3, we get

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Remarks :

- If A and B are mutually exclusive, i.e., $A \cap B = \phi$, then according to theorem 1, we have $P(A \cap B) = 0$. The addition rule, in this case, becomes $P(A \cup B) = P(A) + P(B)$, which is in conformity with axiom III.

2. The event $A \cup B$ denotes the occurrence of either A or B or both. Alternatively, it implies the occurrence of at least one of the two events.
3. The event $A \cap B$ is a compound event that denotes the simultaneous occurrence of the two events.
4. Alternatively, the event $A \cup B$ is also denoted by $A + B$ and the event $A \cap B$ by AB .

Corollaries:

1. From the Venn diagram, we can write $P(A \cup B) = 1 - P(\bar{A} \cap \bar{B})$, where $P(\bar{A} \cap \bar{B})$ is the probability that none of the events A and B occur simultaneously.
2.
$$\begin{aligned} P(\text{exactly one of } A \text{ and } B \text{ occurs}) &= P[(A \cap \bar{B}) \cup (\bar{A} \cap B)] \\ &= P(A \cap \bar{B}) + P(\bar{A} \cap B) \quad [\text{Since } (A \cap \bar{B}) \cup (\bar{A} \cap B) = \phi] \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B) \quad (\text{using theorem 3}) \\ &= P(A \cup B) - P(A \cap B) \quad (\text{using theorem 4}) \end{aligned}$$
3. The addition theorem can be generalised for more than two events. If A, B and C are three events of a sample space S, then the probability of occurrence of at least one of them is given by

$$\begin{aligned} P(A \cup B \cup C) &= P[A \cup (B \cup C)] = P(A) + P(B \cup C) - P[A \cap (B \cup C)] \\ &= P(A) + P(B \cup C) - P[(A \cap B) \cup (A \cap C)] \end{aligned}$$

Applying theorem 4 on the second and third term, we get

$$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \quad \dots (1)$$

Alternatively, the probability of occurrence of at least one of the three events can also be written as

$$P(A \cup B \cup C) = 1 - P(\bar{A} \cap \bar{B} \cap \bar{C}) \quad \dots (2)$$

If A, B and C are mutually exclusive, then equation (1) can be written as

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) \quad \dots (3)$$

If A_1, A_2, \dots, A_n are n events of a sample space S, the respective equations (1), (2) and (3) can be modified as

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n) &= \sum P(A_i) - \sum \sum P(A_i \cap A_j) + \sum \sum \sum P(A_i \cap A_j \cap A_k) \\ &\quad + (-1)^n P(A_1 \cap A_2 \cap \dots \cap A_n) \quad (i \neq j \neq k, \text{ etc.}) \end{aligned} \quad \dots (4)$$

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n) \quad \dots (5)$$

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i) \quad \dots (6)$$

(if the events are mutually exclusive)

4. The probability of occurrence of at least two of the three events can be written as

$$P[(A \cap B) \cup (B \cap C) \cup (A \cap C)] = P(A \cap B) + P(B \cap C) + P(A \cap C) -$$

$$3P(A \cap B \cap C) + P(A \cap B \cap C)$$

$$= P(A \cap B) + P(B \cap C) + P(A \cap C) - 2P(A \cap B \cap C)$$

5. The probability of occurrence of exactly two of the three events can be written as

$$P[(A \cap B \cap \bar{C}) \cup (A \cap \bar{B} \cap C) \cup (\bar{A} \cap B \cap C)] = P[(A \cap B) \cup (B \cap C) \cup (A \cap C)]$$

$$= P(A \cap B \cap C) \text{ (using corollary 2)}$$

$$= P(A \cap B) + P(B \cap C) + P(A \cap C) - 3P(A \cap B \cap C) \text{ (using corollary 4)}$$

6. The probability of occurrence of exactly one of the three events can be written as

$$P[(A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C)] = P(\text{at least one of the three events occur})$$

- P(at least two of the three events occur).

$$= P(A) + P(B) + P(C) - 2P(A \cap B) - 3P(B \cap C) - 2P(A \cap C) + 3P(A \cap B \cap C).$$

Example 23: In a group of 1,000 persons, there are 650 who can speak Hindi, 400 can speak English and 150 can speak both Hindi and English. If a person is selected at random, what is the probability that he speaks (i) Hindi only, (ii) English only, (iii) only one of the two languages, (iv) at least one of the two languages?

Solution: Let A denote the event that a person selected at random speaks Hindi and B denotes the event that he speaks English.

Thus, we have $n(A) = 650$, $n(B) = 400$, $n(A \cap B) = 150$ and $n(S) = 1000$, where $n(A)$, $n(B)$, etc. denote the number of persons belonging to the respective event.

- (i) The probability that a person selected at random speaks Hindi only, is given by

$$P(A \cap \bar{B}) = \frac{n(A)}{n(S)} - \frac{n(A \cap B)}{n(S)} = \frac{650}{1000} - \frac{150}{1000} = \frac{1}{2}$$

- (ii) The probability that a person selected at random speaks English only, is given

$$\text{by } P(\bar{A} \cap B) = \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)} = \frac{400}{1000} - \frac{150}{1000} = \frac{1}{4}$$

- (iii) The probability that a person selected at random speaks only one of the languages, is given by

$$P[(A \cap \bar{B}) \cup (\bar{A} \cap B)] = P(A) + P(B) - 2P(A \cap B) \text{ (see corollary 2)}$$

$$= \frac{n(A) + n(B) - 2n(A \cap B)}{n(S)} = \frac{650 + 400 - 300}{1000} = \frac{3}{4}$$

- (iv) The probability that a person selected at random speaks at least one of the languages, is given by

$$P(A \cup B) = \frac{650 + 400 - 150}{1000} = \frac{9}{10}$$

Alternative Method: The above probabilities can easily be computed by the following nine-square table :

	B	\bar{B}	Total
A	150	500	650
\bar{A}	250	100	350
Total	400	600	1000

From the above table, we can write

$$(i) P(A \cap \bar{B}) = \frac{500}{1000} = \frac{1}{2}$$

$$(ii) P(\bar{A} \cap B) = \frac{250}{1000} = \frac{1}{4}$$

$$(iii) P[(A \cap \bar{B}) \cup (\bar{A} \cap B)] = \frac{500 + 250}{1000} = \frac{3}{4}$$

$$(iv) P(A \cup B) = \frac{150 + 500 + 250}{1000} = \frac{9}{10}$$

This can, alternatively, be written as $P(A \cup B) = 1 - P(\bar{A} \cap \bar{B}) = 1 - \frac{100}{1000} = \frac{9}{10}$.

Example 24: What is the probability of drawing a black card or a king from a well-shuffled pack of playing cards?

Solution: There are 52 cards in a pack, $\therefore n(S) = 52$.

Let A be the event that the drawn card is black and B be the event that it is a king. We have to find $P(A \cup B)$.

Since there are 26 black cards, 4 kings and two black kings in a pack, we have

$$n(A) = 26, n(B) = 4 \text{ and } n(A \cup B) = 26 + 4 - 2 = 28 \text{ Thus, } P(A \cup B) = \frac{28}{52} = \frac{7}{13}$$

Alternative Method: The given information can be written in the form of the following table:

	B	B	Total
A	2	24	26
A	2	24	26
Total	4	48	52

From the above, we can write

$$P(A \cup B) = 1 - P(\bar{A} \cap \bar{B}) = 1 - \frac{48}{52} = \frac{7}{13}$$

Example 25: A pair of unbiased dice is thrown. Find the probability that (i) the sum of spots is either 5 or 10, (ii) either there is a doublet or a sum less than 6.

Solution: Since the first die can be thrown in 6 ways and the second also in 6 ways, therefore, both can be thrown in 36 ways (fundamental principle of counting). Since both the dice are given to be unbiased, 36 elementary outcomes are equally likely.

(i) Let A be the event that the sum of spots is 5 and B be the event that their sum is 10. Thus, we can write

$$A = \{(1, 4), (2, 3), (3, 2), (4, 1)\} \text{ and } B = \{(4, 6), (5, 5), (6, 4)\}$$

We note that $(A \cap B) = \emptyset$, i.e. A and B are mutually exclusive.

$$\therefore \text{By addition theorem, we have } P(A \cup B) = P(A) + P(B) = \frac{4}{36} + \frac{3}{36} = \frac{7}{36}.$$

(ii) Let C be the event that there is a doublet and D be the event that the sum is less than 6. Thus, we can write

$$C = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\} \text{ and}$$

$$D = \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (4, 1)\}$$

Further, $(C \cap D) = \{(1, 1), (2, 2)\}$

$$\text{By addition theorem, we have } P(C \cup D) = \frac{6}{36} + \frac{10}{36} - \frac{2}{36} = \frac{7}{18}.$$

- (i) It is given that $n(A) = 4$, $n(B) = 3$ and $n(S) = 36$. Also $n(A \cap B) = 0$. Thus, the corresponding nine-square table can be written as follows :

	B	\bar{B}	Total
A	0	4	4
\bar{A}	3	29	32
Total	3	33	36

From the above table, we have $P(A \cup B) = 1 - \frac{29}{36} = \frac{7}{36}$.

- (ii) Here $n(C) = 6$, $n(D) = 10$, $n(C \cap D) = 2$ and $n(S) = 36$. Thus, we have

	C	\bar{C}	Total
D	2	8	10
\bar{D}	4	22	26
Total	6	30	36

Thus, $P(C \cup D) = 1 - P(\bar{C} \cap \bar{D}) = 1 - \frac{22}{36} = \frac{7}{18}$.

Example 26: Two unbiased coins are tossed. Let A_1 be the event that the first coin shows a tail and A_2 be the event that the second coin shows a head. Are A_1 and A_2 mutually exclusive? Obtain $P(A_1 \cap A_2)$ and $P(A_1 \cup A_2)$. Further, let A_1 be the event that both coins show heads and A_2 be the event that both show tails. Are A_1 and A_2 mutually exclusive? Find $P(A_1 \cap A_2)$ and $P(A_1 \cup A_2)$.

Solution: The sample space of the experiment is $S = \{(H, H), (H, T), (T, H), (T, T)\}$

- (i) $A_1 = \{(T, H), (T, T)\}$ and $A_2 = \{(H, H), (T, H)\}$

Also $(A_1 \cap A_2) = \{(T, H)\}$, Since $A_1 \cap A_2 \neq \emptyset$, A_1 and A_2 are not mutually exclusive. Further, the coins are given to be unbiased, therefore, all the elementary events are equally likely.

$$\therefore P(A_1) = \frac{2}{4} = \frac{1}{2}, P(A_2) = \frac{2}{4} = \frac{1}{2}, P(A_1 \cap A_2) = \frac{1}{4}$$

Thus, $P(A_1 \cup A_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$.

- (ii) When both the coins show heads; $A_1 = \{(H, H)\}$

When both the coins show tails; $A_2 = \{(T, T)\}$

Here $A_1 \cap A_2 = \emptyset$, $\therefore A_1$ and A_2 are mutually exclusive.

$$\text{Thus, } P(A_1 \cup A_2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Alternatively, the problem can also be attempted by making the following nine-square tables for the two cases :

(i)	A_2	\bar{A}_2	Total	(ii)	A_2	\bar{A}_2	Total
A_1	1	1	2	A_2	0	1	1
\bar{A}_1	1	1	2	\bar{A}_2	1	2	3
Total	2	2	4	Total	1	3	4

Theorem 5: Multiplication or Compound Probability Theorem: A compound event is the result of the simultaneous occurrence of two or more events. For convenience, we assume that there are two events, however, the results can be easily generalised. The probability of the compound event would depend upon whether the events are independent or not. Thus, we shall discuss two theorems; (a) *Conditional Probability Theorem*, and (b) *Multiplicative Theorem for Independent Events*.

(a) **Conditional Probability Theorem:** For any two events A and B in a sample space S, the probability of their simultaneous occurrence, is given by

$$P(A \cap B) = P(A)P(B/A)$$

or equivalently $= P(B)P(A/B)$

Here, $P(B/A)$ is the conditional probability of B given that A has already occurred. Similar interpretation can be given to the term $P(A/B)$.

Proof: Let all the outcomes of the random experiment be equally likely. Therefore,

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{\text{no. of elements in } (A \cap B)}{\text{no. of elements in sample space}}$$

For the event B/A, the sample space is the set of elements in A and out of these the number of cases favourable to B is given by $n(A \cap B)$.

$$\therefore P(B/A) = \frac{n(A \cap B)}{n(A)}.$$

If we multiply the numerator and denominator of the above expression by

$$n(S), \text{ we get } P(B/A) = \frac{n(A \cap B)}{n(A)} \times \frac{n(S)}{n(S)} = \frac{P(A \cap B)}{P(A)}$$

$$\text{or } P(A \cap B) = P(A).P(B/A).$$

The other result can also be shown in a similar way.

Note: To avoid mathematical complications, we have assumed that the elementary events are equally likely. However, the above results will hold true even for the cases where the elementary events are not equally likely.

(b) **Multiplicative Theorem for Independent Events:** If A and B are independent, the probability of their simultaneous occurrence is given by $P(A \cap B) = P(A).P(B)$.

Proof: We can write $A = (A \cap B) \cup (A \cap \bar{B})$.

Since $(A \cap B)$ and $(A \cap \bar{B})$ are mutually exclusive, we have

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) \quad (\text{by axiom III})$$

$$= P(B).P(A/B) + P(\bar{B}).P(A/\bar{B})$$

If A and B are independent, then proportion of A's in B is equal to proportion of A's in \bar{B} 's, i.e., $P(A/B) = P(A/\bar{B})$.

Thus, the above equation can be written as

$$n(B) = \frac{600 \times 30}{100} + \frac{400 \times 5}{100} = 200$$

Substituting this value in the formula of conditional probability theorem, we get

Probability

$$P(A \cap B) = P(A) \cdot P(B).$$

Corollaries:

1. (i) If A and B are mutually exclusive and $P(A) \cdot P(B) > 0$, then they cannot be independent since $P(A \cap B) = 0$.
- (ii) If A and B are independent and $P(A) \cdot P(B) > 0$, then they cannot be mutually exclusive since $P(A \cap B) > 0$.
2. Generalisation of Multiplicative Theorem :

If A, B and C are three events, then

$$P(A \cap B \cap C) = P(A) \cdot P(B / A) \cdot P[C / (A \cap B)]$$

Similarly, for n events A_1, A_2, \dots, A_n , we can write

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_1) \cdot P(A_2 / A_1) \cdot P[A_3 / (A_1 \cap A_2)] \\ &\dots P[A_n / (A_1 \cap A_2 \cap \dots \cap A_{n-1})] \end{aligned}$$

Further, if A_1, A_2, \dots, A_n are independent, we have

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \dots P(A_n).$$

3. If A and B are independent, then A and \bar{B} , \bar{A} and B, \bar{A} and \bar{B} are also independent.

We can write $P(A \cap \bar{B}) = P(A) - P(A \cap B)$ (by theorem 3)

$P(A) - P(A) \cdot P(B) = P(A)[1 - P(B)] = P(A) \cdot P(\bar{B})$, which shows that A and \bar{B} are independent. The other results can also be shown in a similar way.

4. The probability of occurrence of at least one of the events A_1, A_2, \dots, A_n , is given by $P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n)$.

If A_1, A_2, \dots, A_n are independent then their compliments will also be independent, therefore, the above result can be modified as

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - P(\bar{A}_1) \cdot P(\bar{A}_2) \dots P(\bar{A}_n).$$

Pair-wise and Mutual Independence

Three events A, B and C are said to be mutually independent if the following conditions are simultaneously satisfied :

$$P(A \cap B) = P(A) \cdot P(B), P(B \cap C) = P(B) \cdot P(C), P(A \cap C) = P(A) \cdot P(C)$$

$$\text{and } P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C).$$

If the last condition is not satisfied, the events are said to be pair-wise independent.

From the above we note that mutually independent events will always be pair-wise independent but not vice-versa.

Example 27: Among 1,000 applicants for admission to M.A. economics course in a University, 600 were economics graduates and 400 were non-economics graduates; 30% of economics graduate applicants and 5% of non-economics graduate applicants obtained admission. If an applicant selected at random is found to have been given admission, what is the probability that he/she is an economics graduate?

Solution: Let A be the event that the applicant selected at random is an economics graduate and B be the event that he/she is given admission.

We are given $n(S) = 1000$, $n(A) = 600$, $n(\bar{A}) = 400$

Also, $n(B) = \frac{600 \times 30}{100} + \frac{400 \times 5}{100} = 200$ and $n(A \cap B) = \frac{600 \times 30}{100} = 180$

Thus, the required probability is given by $P(A/B) = \frac{n(A \cap B)}{n(B)} = \frac{180}{200} = \frac{9}{10}$

Alternative Method: Writing the given information in a nine-square table, we have :

	B	\bar{B}	Total
A	180	420	600
\bar{A}	20	380	400
Total	200	800	1000

From the above table we can write $P(A/B) = \frac{180}{200} = \frac{9}{10}$

Example 28: A bag contains 2 black and 3 white balls. Two balls are drawn at random one after the other without replacement. Obtain the probability that (a) Second ball is black given that the first is white, (b) First ball is white given that the second is black.

Solution: First ball can be drawn in any one of the 5 ways and then a second ball can be drawn in any one of the 4 ways. Therefore, two balls can be drawn in $5 \times 4 = 20$ ways. Thus, $n(S) = 20$.

(a) Let A_1 be the event that first ball is white and A_2 be the event that second is black. We want to find $P(A_2 / A_1)$.

First white ball can be drawn in any of the 3 ways and then a second ball can be drawn in any of the 4 ways, $\therefore n(A_1) = 3 \times 4 = 12$.

Further, first white ball can be drawn in any of the 3 ways and then a black ball can be drawn in any of the 2 ways, $\therefore n(A_1 \cap A_2) = 3 \times 2 = 6$.

Thus, $P(A_2 / A_1) = \frac{n(A_1 \cap A_2)}{n(A_1)} = \frac{6}{12} = \frac{1}{2}$.

(b) Here we have to find $P(A_1 / A_2)$.

The second black ball can be drawn in the following two mutually exclusive ways:

- (i) First ball is white and second is black or
- (ii) both the balls are black.

Thus, $n(A_2) = 3 \times 2 + 2 \times 1 = 8$, $\therefore P(A_1 / A_2) = \frac{n(A_1 \cap A_2)}{n(A_2)} = \frac{6}{8} = \frac{3}{4}$.

Alternative Method: The given problem can be summarised into the following nine-square table:

	B	\bar{B}	Total
A	6	6	12
\bar{A}	2	6	8
Total	8	12	20

The required probabilities can be directly written from the above table.

Example 29: Two unbiased dice are tossed. Let w denote the number on the first die and r denote the number on the second die. Let A be the event that $w + r \leq 4$ and B be the event that $w + r \leq 3$. Are A and B independent?

Solution: The sample space of this experiment consists of 36 elements, i.e., $n(S) = 36$. Also, $A = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)\}$ and $B = \{(1, 1), (1, 2), (2, 1)\}$.

From the above, we can write

$$P(A) = \frac{6}{36} = \frac{1}{6}, P(B) = \frac{3}{36} = \frac{1}{12}$$

$$\text{Also } (A \cap B) = \{(1, 1), (1, 2), (2, 1)\} \therefore P(A \cap B) = \frac{3}{36} = \frac{1}{12}$$

Since $P(A \cap B) \neq P(A)P(B)$, A and B are not independent.

Example 30: It is known that 40% of the students in a certain college are girls and 50% of the students are above the median height. If $2/3$ of the boys are above median height, what is the probability that a randomly selected student who is below the median height is a girl?

Solution: Let A be the event that a randomly selected student is a girl and B be the event that he/she is above median height. The given information can be summarised into the following table :

	B	\bar{B}	Total
A	10	30	40
\bar{A}	40	20	60
Total	50	50	100

From the above table, we can write $P(A/\bar{B}) = \frac{30}{50} = 0.6$.

Example 31: A problem in statistics is given to three students A , B and C , whose chances of solving it independently are $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ respectively. Find the probability that

- (a) the problem is solved.
- (b) at least two of them are able to solve the problem.
- (c) exactly two of them are able to solve the problem.
- (d) exactly one of them is able to solve the problem.

Solution: Let A be the event that student A solves the problem. Similarly, we can define the events B and C . Further, A , B and C are given to be independent.

- (a) The problem is solved if at least one of them is able to solve it. This probability is

$$\text{given by } P(A \cup B \cup C) = 1 - P(\bar{A}) \cdot P(\bar{B}) \cdot P(\bar{C}) = 1 - \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} = \frac{3}{4}$$

- (b) Here we have to find $P[(A \cap B) \cup (B \cap C) \cup (A \cap C)]$

$$P[(A \cap B) \cup (B \cap C) \cup (A \cap C)] = P(A)P(B) + P(B)P(C) + P(A)P(C) \\ - 2P(A)P(B)P(C)$$

$$= \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{4} - 2 \cdot \frac{1}{2} \times \frac{1}{3} \times \frac{1}{4} = \frac{7}{24}$$

(c) The required probability is given by $P[(A \cap B \cap \bar{C}) \cup (\bar{A} \cap B \cap C) \cup (\bar{A} \cap \bar{B} \cap C)]$

$$= P(A).P(B) + P(\bar{B}).P(C) + P(\bar{A}).P(\bar{B}).P(C)$$

$$= \frac{1}{6} + \frac{1}{12} + \frac{1}{8} - \frac{1}{8} = \frac{1}{4}.$$

(d) The required probability is given by $P[(A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C)]$

$$= P(A) + P(\bar{B}) + P(C) - 2P(A).P(\bar{B}) - 2P(\bar{B}).P(C) - 2P(A).P(C)$$

$$= \frac{1}{2} + \frac{1}{3} + \frac{1}{4} - \frac{1}{3} - \frac{1}{6} - \frac{1}{4} + \frac{1}{8} = \frac{11}{24}.$$

Note that the formulae used in (a), (b), (c) and (d) above are the modified forms of corollaries (following theorem 4) 3, 4, 5 and 6 respectively.

Example 32: A bag contains 2 red and 1 black ball and another bag contains 2 red and 2 black balls. One ball is selected at random from each bag. Find the probability of drawing (a) at least a red ball, (b) a black ball from the second bag given that ball from the first is red; (c) show that the event of drawing a red ball from the first bag and the event of drawing a red ball from the second bag are independent.

Solution: Let A_1 be the event of drawing a red ball from the first bag and A_2 be the event of drawing a red ball from the second bag. Thus, we can write:

$$n(A_1 \cap A_2) = 2 \times 2 = 4, \quad n(A_1 \cap \bar{A}_2) = 2 \times 2 = 4,$$

$$n(\bar{A}_1 \cap A_2) = 1 \times 2 = 2, \quad n(\bar{A}_1 \cap \bar{A}_2) = 1 \times 2 = 2$$

$$\text{Also, } n(S) = n(A_1 \cap A_2) + n(A_1 \cap \bar{A}_2) + n(\bar{A}_1 \cap A_2) + n(\bar{A}_1 \cap \bar{A}_2) = 12$$

Writing the given information in the form of a nine-square table, we get

	A_2	\bar{A}_2	Total
A_1	4	4	8
\bar{A}_1	2	2	4
Total	6	6	12

(a) The probability of drawing at least a red ball is given by

$$P(A_1 \cup A_2) = 1 - \frac{n(\bar{A}_1 \cap \bar{A}_2)}{n(S)} = 1 - \frac{2}{12} = \frac{5}{6}$$

(b) We have to find $P(\bar{A}_2 / A_1)$

$$P(\bar{A}_2 / A_1) = \frac{n(A_1 \cap \bar{A}_2)}{n(A_1)} = \frac{4}{8} = \frac{1}{2}$$

(c) A_1 and A_2 will be independent if $P(A_1 \cap A_2) = P(A_1).P(A_2)$

$$\text{Now } P(A_1 \cap A_2) = \frac{n(A_1 \cap A_2)}{n(S)} = \frac{4}{12} = \frac{1}{3}$$

$$P(A_1).P(A_2) = \frac{n(A_1)}{n(S)} \cdot \frac{n(A_2)}{n(S)} = \frac{8}{12} \times \frac{6}{12} = \frac{1}{3}$$

Hence, A_1 and A_2 are independent.

Example 33: An urn contains 3 red and 2 white balls. 2 balls are drawn at random. Find the probability that either both of them are red or both are white.

Solution: Let A be the event that both the balls are red and B be the event that both the balls are white. Thus, we can write

$$n(S) = {}^5C_2 = 10, n(A) = {}^3C_2 = 3, n(B) = {}^2C_2 = 1, \text{ also } n(A \cap B) = 0$$

$$\therefore \text{The required probability is } P(A \cup B) = \frac{n(A) + n(B)}{n(S)} = \frac{3+1}{10} = \frac{2}{5}$$

Example 34: A bag contains 10 red and 8 black balls. Two balls are drawn at random. Find the probability that (a) both of them are red, (b) one is red and the other is black.

Solution: Let A be the event that both the balls are red and B be the event that one is red and the other is black.

Two balls can be drawn from 18 balls in ${}^{18}C_2$ equally likely ways.

$$\therefore n(S) = {}^{18}C_2 = \frac{18!}{2!16!} = 153$$

(a) Two red balls can be drawn from 10 red balls in ${}^{10}C_2$ ways.

$$\therefore n(A) = {}^{10}C_2 = \frac{10!}{2!8!} = 45$$

$$\text{Thus, } P(A) = \frac{n(A)}{n(S)} = \frac{45}{153} = \frac{5}{17}$$

(b) One red ball can be drawn in ${}^{10}C_1$ ways and one black ball can be drawn in 8C_1 ways.

$$\therefore n(B) = {}^{10}C_1 \times {}^8C_1 = 10 \times 8 = 80 \text{ Thus, } P(B) = \frac{80}{153}$$

Example 35: Five cards are drawn in succession and without replacement from an ordinary deck of 52 well-shuffled cards :

- (a) What is the probability that there will be no ace among the five cards?
- (b) What is the probability that first three cards are aces and the last two cards are kings?
- (c) What is the probability that only first three cards are aces?
- (d) What is the probability that an ace will appear only on the fifth draw?

Solution:

$$(a) P(\text{there is no ace}) = \frac{48 \times 47 \times 46 \times 45 \times 44}{52 \times 51 \times 50 \times 49 \times 48} = 0.66$$

$$(b) P\left(\begin{array}{l} \text{first three card are aces and} \\ \text{the last two are kings} \end{array}\right) = \frac{4 \times 3 \times 2 \times 4 \times 3}{52 \times 51 \times 50 \times 49 \times 48} = 0.0000009$$

$$(c) P(\text{only first three card are aces}) = \frac{4 \times 3 \times 2 \times 48 \times 47}{52 \times 51 \times 50 \times 49 \times 48} = 0.00017$$

$$(d) P\left(\begin{array}{l} \text{an ace appears only} \\ \text{on the fifth draw} \end{array}\right) = \frac{48 \times 47 \times 46 \times 45 \times 4}{52 \times 51 \times 50 \times 49 \times 48} = 0.059$$

Example 36: Two cards are drawn in succession from a pack of 52 well-shuffled cards. Find the probability that :

- (a) Only first card is a king.
- (b) First card is jack of diamond or a king.
- (c) At least one card is a picture card.
- (d) Not more than one card is a picture card.
- (e) Cards are not of the same suit.
- (f) Second card is not a spade.
- (g) Second card is not a spade given that first is a spade.
- (h) The cards are aces or diamonds or both.

Solution:

- (a) $P(\text{only first card is a king}) = \frac{4 \times 48}{52 \times 51} = \frac{16}{221}$.
- (b) $P\left(\begin{array}{l}\text{first card is a jack of} \\ \text{diamond or a king}\end{array}\right) = \frac{5 \times 51}{52 \times 51} = \frac{5}{52}$.
- (c) $P\left(\begin{array}{l}\text{at least one card is} \\ \text{a picture card}\end{array}\right) = 1 - \frac{40 \times 39}{52 \times 51} = \frac{7}{17}$.
- (d) $P\left(\begin{array}{l}\text{not more than one card} \\ \text{is a picture card}\end{array}\right) = \frac{40 \times 39}{52 \times 51} + \frac{12 \times 40}{52 \times 51} + \frac{40 \times 12}{52 \times 51} = \frac{210}{221}$.
- (e) $P(\text{cards are not of the same suit}) = \frac{52 \times 39}{52 \times 51} = \frac{13}{17}$.
- (f) $P(\text{second card is not a spade}) = \frac{13 \times 39}{52 \times 51} + \frac{39 \times 38}{52 \times 51} = \frac{3}{4}$.
- (g) $P\left(\begin{array}{l}\text{second card is not a spade} \\ \text{given that first is spade}\end{array}\right) = \frac{39}{51} = \frac{13}{17}$.
- (h) $P\left(\begin{array}{l}\text{the cards are aces or} \\ \text{diamonds or both}\end{array}\right) = \frac{16 \times 15}{52 \times 51} = \frac{20}{221}$.

Example 37: The odds are 9 : 7 against a person A, who is now 35 years of age, living till he is 65 and 3 : 2 against a person B, now 45 years of age, living till he is 75. Find the chance that at least one of these persons will be alive 30 years hence.

Solution:

Note: If a is the number of cases favourable to an event A and α is the number of cases favourable to its compliment event ($a + \alpha = n$), then odds in favour of A are $a : \alpha$ and odds against A are $\alpha : a$.

$$\text{Obviously } P(A) = \frac{a}{a + \alpha} \text{ and } P(\bar{A}) = \frac{\alpha}{a + \alpha}.$$

Let A be the event that person A will be alive 30 years hence and B be the event that person B will be alive 30 years hence.

$$\therefore P(A) = \frac{7}{9+7} = \frac{7}{16} \text{ and } P(B) = \frac{2}{3+2} = \frac{2}{5}$$

We have to find $P(A \cup B)$. Note that A and B are independent.

$$\therefore P(A \cup B) = \frac{7}{16} + \frac{2}{5} - \frac{7}{16} \times \frac{2}{5} = \frac{53}{80}$$

$$P(A \cup B) = 1 - \frac{9}{16} \times \frac{3}{5} = \frac{53}{80}$$

Example 38: If A and B are two events such that $P(A) = \frac{2}{3}$, $P(\bar{A} \cap B) = \frac{1}{6}$ and

$$P(A \cap B) = \frac{1}{3}, \text{ find } P(B), P(A \cup B), P(A/B), P(B/A), P(\bar{A} \cup B), P(\bar{A} \cap \bar{B}) \text{ and } P(\bar{B}).$$

Also examine whether the events A and B are : (a) Equally likely, (b) Exhaustive, (c) Mutually exclusive and (d) Independent.

Solution: The probabilities of various events are obtained as follows :

$$P(B) = P(\bar{A} \cap B) + P(A \cap B) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

$$P(A \cup B) = \frac{2}{3} + \frac{1}{2} - \frac{1}{3} = \frac{5}{6}$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3} \times \frac{2}{1} = \frac{2}{3}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1}{3} \times \frac{3}{2} = \frac{1}{2}$$

$$P(\bar{A} \cup B) = P(\bar{A}) + P(B) - P(\bar{A} \cap B) = \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{2}{3}$$

$$P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B) = 1 - \frac{5}{6} = \frac{1}{6}$$

$$P(\bar{B}) = 1 - P(B) = 1 - \frac{1}{2} = \frac{1}{2}$$

- (a) Since $P(A) \neq P(B)$, A and B are not equally likely events.
- (b) Since $P(A \cup B) \neq 1$, A and B are not exhaustive events.
- (c) Since $P(A \cap B) \neq 0$, A and B are not mutually exclusive.
- (d) Since $P(A)P(B) = P(A \cap B)$, A and B are independent events.

Example 39: Two players A and B toss an unbiased die alternatively. He who first throws a six wins the game. If A begins, what is the probability that B wins the game?

Solution: Let A_i and B_i be the respective events that A and B throw a six in i^{th} toss, $i = 1, 2, \dots$. B will win the game if any one of the following mutually exclusive events occur: $\bar{A}_1 B_1$ or $\bar{A}_1 \bar{B}_1 \bar{A}_2 B_2$ or $\bar{A}_1 \bar{B}_1 \bar{A}_2 \bar{B}_2 \bar{A}_3 B_3$, etc.

$$\text{Thus, } P(\text{B wins}) = \frac{5}{6} \times \frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + \dots$$

$$= \frac{5}{36} \left[1 + \left(\frac{5}{6} \right)^2 + \left(\frac{5}{6} \right)^4 + \dots \right] = \frac{5}{36} \times \frac{1}{1 - \left(\frac{5}{6} \right)^2} = \frac{5}{11}$$

Example 40: A bag contains 5 red and 3 black balls and second bag contains 4 red and 5 black balls.

- (a) If one ball is selected at random from each bag, what is the probability that both of them are of same colour?
- (b) If a bag is selected at random and two balls are drawn from it, what is the probability that they are of (i) same colour, (ii) different colours?

Solution:

$$(a) \text{ Required Probability} = \left[\begin{array}{l} \text{Probability that ball} \\ \text{from both bags are red} \end{array} \right] + \left[\begin{array}{l} \text{Probability that balls} \\ \text{from both bags are black} \end{array} \right]$$

$$= \frac{5}{8} \times \frac{4}{9} + \frac{3}{8} \times \frac{5}{9} = \frac{35}{72}$$

- (b) Let A be the event that first bag is drawn so that \bar{A} denotes the event that second bag is drawn. Since the two events are equally likely, mutually exclusive and exhaustive, we have $P(A) = P(\bar{A}) = \frac{1}{2}$.

- (i) Let R be the event that two drawn balls are red and B be the event that they are black. The required probability is given by

$$= P(A)[P(R/A) + P(B/A)] + P(\bar{A})[P(R/\bar{A}) + P(B/\bar{A})]$$

$$= \frac{1}{2} \left[\frac{{}^5C_2 + {}^3C_2}{{}^8C_2} \right] + \frac{1}{2} \left[\frac{{}^4C_2 + {}^5C_2}{{}^9C_2} \right] = \frac{1}{2} \left[\frac{10+3}{28} \right] + \frac{1}{2} \left[\frac{6+10}{36} \right] = \frac{229}{504}$$

- (ii) Let C denote the event that the drawn balls are of different colours. The required probability is given by

$$P(C) = P(A)P(C/A) + P(\bar{A})P(C/\bar{A})$$

$$= \frac{1}{2} \left[\frac{5 \times 3}{{}^8C_2} \right] + \frac{1}{2} \left[\frac{4 \times 5}{{}^9C_2} \right] = \frac{1}{2} \left[\frac{15}{28} + \frac{20}{36} \right] = \frac{275}{504}$$

Example 41: There are two urns U_1 and U_2 . U_1 contains 9 white and 4 red balls and U_2 contains 3 white and 6 red balls. Two balls are transferred from U_1 to U_2 and then a ball is drawn from U_2 . What is the probability that it is a white ball?

Solution: Let A be the event that the two transferred balls are white, B be the event that they are red and C be the event that one is white and the other is red. Further, let W be the event that a white ball is drawn from U_2 . The event W can occur with any one of the mutually exclusive events A, B and C.

$$P(W) = P(A).P(W/A) + P(B)P(W/B) + P(C)P(W/C)$$

$$= \frac{{}^9C_2}{{}^{13}C_2} \times \frac{5}{11} + \frac{{}^4C_2}{{}^{13}C_2} \times \frac{3}{11} + \frac{9 \times 4}{{}^{13}C_2} \times \frac{4}{11} = \frac{57}{143}$$

Example 42: A bag contains tickets numbered as 112, 121, 211 and 222. One ticket is drawn at random from the bag. Let E_i ($i = 1, 2, 3$) be the event that i th digit on the ticket is

2. Discuss the independence of E_1 , E_2 and E_3 .

Solution: The event E_1 occurs if the number on the drawn ticket 211 or 222, therefore,

$$P(E_1) = \frac{1}{2}. \text{ Similarly } P(E_2) = \frac{1}{2} \text{ and } P(E_3) = \frac{1}{2}.$$

$$\text{Now } P(E_i \cap E_j) = \frac{1}{4} \text{ (i, j = 1, 2, 3 and i ≠ j).}$$

Since $P(E_i \cap E_j) = P(E_i)P(E_j)$ for $i \neq j$, therefore E_1, E_2 and E_3 are pair-wise independent.

Further, $P(E_1 \cap E_2 \cap E_3) = \frac{1}{4} \neq P(E_1).P(E_2).P(E_3)$, therefore, E_1, E_2 and E_3 are not mutually independent.

Example 43: Probability that an electric bulb will last for 150 days or more is 0.7 and that it will last at the most 160 days is 0.8. Find the probability that it will last between 150 to 160 days.

Solution: Let A be the event that the bulb will last for 150 days or more and B be the event that it will last at the most 160 days. It is given that $P(A) = 0.7$ and $P(B) = 0.8$.

The event $A \cup B$ is a certain event because at least one of A or B is bound to occur.

Thus, $P(A \cup B) = 1$. We have to find $P(A \cap B)$. This probability is given by

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.7 + 0.8 - 1.0 = 0.5$$

Example 44: The odds that A speaks the truth are 2 : 3 and the odds that B speaks the truth are 4 : 5. In what percentage of cases they are likely to contradict each other on an identical point?

Solution: Let A and B denote the respective events that A and B speak truth. It is given

$$\text{that } P(A) = \frac{2}{5} \text{ and } P(B) = \frac{4}{9}.$$

The event that they contradict each other on an identical point is given by $(A \cap \bar{B}) \cup (\bar{A} \cap B)$,

where $(A \cap \bar{B})$ and $(\bar{A} \cap B)$ are mutually exclusive. Also A and B are independent events.

Thus, we have

$$\begin{aligned} P[(A \cap \bar{B}) \cup (\bar{A} \cap B)] &= P(A \cap \bar{B}) + P(\bar{A} \cap B) = P(A).P(\bar{B}) + P(\bar{A}).P(B) \\ &= \frac{2}{5} \times \frac{5}{9} + \frac{3}{5} \times \frac{4}{9} = \frac{22}{45} = 0.49 \end{aligned}$$

Hence, A and B are likely to contradict each other in 49% of the cases.

Example 45: The probability that a student A solves a mathematics problem is $\frac{2}{5}$ and

the probability that a student B solves it is $\frac{2}{3}$. What is the probability that (a) the problem is not solved, (b) the problem is solved, (c) Both A and B, working independently of each other, solve the problem?

Solution: Let A and B be the respective events that students A and B solve the problem. We note that A and B are independent events.

$$(a) P(\bar{A} \cap \bar{B}) = P(\bar{A}).P(\bar{B}) = \frac{3}{5} \times \frac{1}{3} = \frac{1}{5}$$

$$(b) P(A \cup B) = 1 - P(\bar{A} \cap \bar{B}) = 1 - \frac{1}{5} = \frac{4}{5}$$

$$(c) P(A \cap B) = P(A)P(B) = \frac{2}{5} \times \frac{2}{3} = \frac{4}{15}$$

Example 46: A bag contains 8 red and 5 white balls. Two successive drawings of 3 balls each are made such that (i) balls are replaced before the second trial, (ii) balls are not replaced before the second trial. Find the probability that the first drawing will give 3 white and the second 3 red balls.

Solution: Let A be the event that all the 3 balls obtained at the first draw are white and B be the event that all the 3 balls obtained at the second draw are red.

(a) When balls are replaced before the second draw, we have

$$P(A) = \frac{^5C_3}{^{13}C_3} = \frac{5}{143} \text{ and } P(B) = \frac{^8C_3}{^{13}C_3} = \frac{28}{143}$$

The required probability is given by $P(A \cap B)$, where A and B are independent. Thus, we have

$$P(A \cap B) = P(A).P(B) = \frac{5}{143} \times \frac{28}{143} = \frac{140}{20449}$$

(b) When the balls are not replaced before the second draw

$$\text{We have } P(B/A) = \frac{^8C_3}{^{10}C_3} = \frac{7}{15}. \text{ Thus, we have}$$

$$P(A \cap B) = P(A).P(B/A) = \frac{5}{143} \times \frac{7}{15} = \frac{7}{429}$$

Example 47: Computers A and B are to be marketed. A salesman who is assigned the job of finding customers for them has 60% and 40% chances respectively of succeeding in case of computer A and B. The two computers can be sold independently. Given that the salesman is able to sell at least one computer, what is the probability that computer A has been sold?

Solution: Let A be the event that the salesman is able to sell computer A and B be the event that he is able to sell computer B. It is given that $P(A) = 0.6$ and $P(B) = 0.4$. The probability that the salesman is able to sell at least one computer, is given by

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A).P(B)$$

(note that A and B are given to be independent)

$$= 0.6 + 0.4 - 0.6 \times 0.4 = 0.76$$

Now the required probability, the probability that computer A is sold given that the salesman is able to sell at least one computer, is given by

$$P(A/A \cup B) = \frac{0.60}{0.76} = 0.789$$

Example 48: Two men M_1 and M_2 and three women W_1 , W_2 and W_3 , in a big industrial firm, are trying for promotion to a single post which falls vacant. Those of the same sex have equal probabilities of getting promotion but each man is twice as likely to get the promotion as any women.

- (a) Find the probability that a woman gets the promotion.
 (b) If M_2 and W_2 are husband and wife, find the probability that one of them gets the promotion.

Solution: Let p be the probability that a woman gets the promotion, therefore $2p$ will be the probability that a man gets the promotion. Thus, we can write, $P(M_1) = P(M_2) = 2p$ and $P(W_1) = P(W_2) = P(W_3) = p$, where $P(M_i)$ denotes the probability that i th man gets the promotion ($i = 1, 2$) and $P(W_j)$ denotes the probability that j th woman gets the promotion.

Since the post is to be given only to one of the five persons, the events M_1, M_2, W_1, W_2 and W_3 are mutually exclusive and exhaustive.

$$\therefore P(M_1 \cup M_2 \cup W_1 \cup W_2 \cup W_3) = P(M_1) + P(M_2) + P(W_1) + P(W_2) + P(W_3) = 1$$

$$\Rightarrow 2p + 2p + p + p + p = 1 \text{ or } p = \frac{1}{7}$$

- (a) The probability that a woman gets the promotion

$$P(W_1 \cup W_2 \cup W_3) = P(W_1) + P(W_2) + P(W_3) = \frac{3}{7}$$

- (b) The probability that M_2 or W_2 gets the promotion

$$P(M_2 \cup W_2) = P(M_2) + P(W_2) = \frac{3}{7}$$

Example 49: An unbiased die is thrown 8 times. What is the probability of getting a six in at least one of the throws?

Solution: Let A_i be the event that a six is obtained in the i th throw ($i = 1, 2, \dots, 8$).

$$\text{Therefore, } P(A_i) = \frac{1}{6}.$$

The event that a six is obtained in at least one of the throws is represented by $(A_1 \cup A_2 \cup \dots \cup A_8)$. Thus, we have

$$P(A_1 \cup A_2 \cup \dots \cup A_8) = 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_8)$$

Since A_1, A_2, \dots, A_8 are independent, we can write

$$P(A_1 \cup A_2 \cup \dots \cup A_8) = 1 - P(\bar{A}_1) \cdot P(\bar{A}_2) \cdot \dots \cdot P(\bar{A}_8) = 1 - \left(\frac{5}{6}\right)^8.$$

Example 50: Two students X and Y are very weak students of mathematics and their chances of solving a problem correctly are 0.11 and 0.14 respectively. If the probability of their making a common mistake is 0.081 and they get the same answer, what is the chance that their answer is correct?

Solution: Let A be the event that both the students get a correct answer, B be the event that both get incorrect answer by making a common mistake and C be the event that both get the same answer. Thus, we have

$$\begin{aligned} P(A \cap C) &= P(X \text{ gets correct answer}) \cdot P(Y \text{ gets correct answer}) \\ &= 0.11 \times 0.14 = 0.0154 \text{ (note that the two events are independent)} \end{aligned}$$

Similarly,

$$\begin{aligned} P(B \cap C) &= P(X \text{ gets incorrect answer}) \times P(Y \text{ gets incorrect answer}) \\ &\quad \times P(X \text{ and } Y \text{ make a common mistake}) \end{aligned}$$

$$= (1 - 0.11)(1 - 0.14) \times 0.081 = 0.062$$

Further, $C = (A \cap C) \cup (B \cap C)$ or $P(C) = P(A \cap C) + P(B \cap C)$, since $(A \cap C)$ and $(B \cap C)$ are mutually exclusive. Thus, we have

$$P(C) = 0.0154 + 0.0620 = 0.0774$$

We have to find the probability that the answers of both the students are correct given that they are same, i.e.,

$$P(A/C) = \frac{P(A \cap C)}{P(C)} = \frac{0.0154}{0.0774} = 0.199$$

Example 51: Given below are the daily wages (in rupees) of six workers of a factory :

77, 105, 91, 100, 90, 83

If two of these workers are selected at random to serve as representatives, what is the probability that at least one will have a wage lower than the average?

Solution: The average wage $\bar{X} = \frac{77+105+91+100+90+83}{6} = 91$

Let A be the event that two workers selected at random have their wages greater than or equal to average wage.

$$\therefore P(A) = \frac{\binom{3}{2}}{\binom{6}{2}} = \frac{1}{5}$$

Thus, the probability that at least one of the workers has a wage less than the average

$$= 1 - \frac{1}{5} = \frac{4}{5}$$

Example 52: There are two groups of subjects one of which consists of 5 science subjects and 3 engineering subjects and the other consists of 3 science subjects and 5 engineering subjects. An unbiased die is cast. If the number 3 or 5 turns up, a subject from the first group is selected at random otherwise a subject is randomly selected from the second group. Find the probability that an engineering subject is selected ultimately.

Solution: Let A be the event that an engineering subject is selected and B be the event that 3 or 5 turns on the die. The given information can be summarised into symbols, as given below :

$$P(A) = \frac{1}{3}, \quad P(A/B) = \frac{3}{8}, \quad \text{and} \quad P(A/\bar{B}) = \frac{5}{8}$$

To find P(A), we write

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap \bar{B}) = P(B) \cdot P(A/B) + P(\bar{B}) \cdot P(A/\bar{B}) \\ &= \frac{1}{3} \times \frac{3}{8} + \frac{2}{3} \times \frac{5}{8} = \frac{13}{24} \end{aligned}$$

Example 53: Find the probability of obtaining two heads in the toss of two unbiased coins when (a) at least one of the coins shows a head, (b) second coin shows a head.

Solution: Let A be the event that both coins show heads, B be the event that at least one coin shows a head and C be the event that second coin shows a head. The sample space and the three events can be written as :

$$S = \{(H, H), (H, T), (T, H), (T, T)\}, \quad A = \{(H, H)\}, \\ B = \{(H, H), (H, T), (T, H)\} \quad \text{and} \quad C = \{(H, H), (T, H)\}.$$

Probability

Further, $A \cap B = \{(H, H)\}$ and $A \cap C = \{(H, H)\}$

Since the coins are given to be unbiased, the elementary events are equally likely, therefore

$$P(A) = \frac{1}{4}, \quad P(B) = \frac{3}{4}, \quad P(C) = \frac{1}{2}, \quad P(A \cap B) = P(A \cap C) = \frac{1}{4}$$

(a) We have to determine $P(A/B)$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{4} \times \frac{4}{3} = \frac{1}{3}$$

(b) We have to determine $P(A/C)$

$$P(A/C) = \frac{P(A \cap C)}{P(C)} = \frac{1}{4} \times \frac{2}{1} = \frac{1}{2}$$

Exercise with Hints:

- What is the probability of drawing two aces at random from a deck of 52 well-shuffled cards?

Hint: Two aces can be drawn from four aces in 4C_2 ways.

- Two cards are drawn at random from a deck of 52 well-shuffled cards. What is the probability that one of them is an ace and the other is a queen?

Hint: Try as in question 1 above.

- What is the probability of getting all the four heads in four throws of an unbiased coin?

Hint: $n(S) = 2^4$.

- What is the probability of getting 5 on each of the two throws of a six faced unbiased die?

Hint: Try as in question 3 above.

- Four cards are drawn at random without replacement from a pack of 52 cards. What is the probability that :

- All of them are aces?
- All of them are of different suits?
- All of them are picture cards or spades or both?

Hint: See example 36.

- Find the probability of throwing an even number from a single throw of a pair of unbiased dice.

Hint: An even number is obtained if both dice show either odd or even numbers.

- A bag contains 50 balls serially numbered from 1 to 50. One ball is drawn at random from the bag. What is the probability that the number on it is a multiple of 3 or 4?

Hint: The number of serial numbers that are multiple of 3 or 4 are integral part of

$$\frac{50}{L.C.M. \text{ of } 3 \text{ and } 4}.$$

8. A bag contains 4 white and 5 red balls. Two balls are drawn in succession at random. What is the probability that (a) both the balls are white, (b) both are red, (c) one of them is red and the other is white?

Hint: See example 34.

9. A bag contains 5 red, 8 white and 3 blue balls. If three balls are drawn at random, find the probability that (a) all the balls are blue, (b) each ball is of different colour, (c) the drawn balls are in the order red, white and blue, (d) none of the balls are white.

Hint: (b) This event is same as that of drawing one ball of each colour.
(c) $n(S) = 16 \times 15 \times 14$.

10. 4 cards are drawn at random from a pack of 52 well-shuffled cards. Find the chance that (i) each card is of a different suit, (ii) they consist of a Jack, Queen, King and an Ace, (iii) they are 4 honours of the same suit.

Hint: Honours of a suit are its Jack, Queen, King and Ace.

11. In how many ways the letters of the following words can be arranged?

MANAGEMENT, ASSESSMENT, COMMITTEE

Hint: See example 13.

12. How many distinct words can be formed from the letters of the word MEERUT? How many of these words start at M and end at T?

Hint: Fixing M and T, determine the number of permutations of remaining letters.

13. In a random arrangement of letters of the word DROUGHT, find the probability that vowels come together.

Hint: See example 15.

14. The letters of the word STUDENT are arranged at random. Find the probability that the word, so formed;

- (a) starts with S,
- (b) starts with S and ends with T,
- (c) the vowels occupy odd positions only,
- (d) the vowels occupy even positions only.

Hint: See examples 14 and 15.

15. How many triangles can be formed by joining 12 points in a plane, given that 7 points are on one line.

Hint: No. of triangles = ${}^{12}C_3 - {}^7C_3$.

16. In a random arrangement of 10 members of a committee, find the probability that there are exactly 3 members sitting between the president and secretary when the arrangement is done (i) in a row, (ii) in a ring.

Hint: Considering 5 members as one, there are 6 members. No. of permutations
(i) $2! \times {}^8C_3 \times 3! \times 6!$, (ii) $2! \times {}^8C_3 \times 3! \times 5!$

17. A six digit number is formed by the digits 5, 9, 0, 7, 1, 3; no digit being repeated. Find the probability that the number formed is (i) divisible by 5, (ii) not divisible by 5.

Hint: 0 cannot come at the sixth place of a six digit number.

18. If 30 blankets are distributed at random among 10 beggars, find the probability that a particular beggar receives 5 blankets.

Hint: A particular beggar can receive 5 blankets in ${}^{30}C_5$ ways and the remaining 9 beggars in 9^{25} ways.

19. A statistical experiment consists of asking 3 housewives, selected at random, if they wash their dishes with brand X detergent. List the elements of the sample space S using the letter Y for 'yes' and N for 'no'. Also list the elements of the event : "The second woman interviewed uses brand X". Find the probability of this event if it is assumed that all the elements of S are equally likely to occur.

Hint: The sample space would consist of eight 3-tuples of the type (Y,Y,Y), etc.

20. n persons are sitting in a row. If two persons are picked up at random, what is the probability that they are sitting adjacent to each other?

Hint: Two adjacent persons can be picked up in $(n - 1)$ ways.

21. A committee of 5 persons is to be formed out of 7 Indians and 5 Japanese. Find the probability that (a) the committee is represented only by the Indians, (b) there are at least two Japanese on the committee, (c) there are at least two Japanese and two Indians on the committee.

Hint: See example 16.

22. 4 letters are placed at random in 4 addressed envelopes. Find the probability that all the letters are not placed in right envelopes.

Hint: The letters can be placed in their respective envelopes in one way.

23. Find the probability that a family with 4 children has (a) 2 boys and 2 girls, (b) no boy, (c) at the most two boys, (d) at least a girl. Assume equal probability for boys and girls.

Hint: (a) The event can occur in 4C_2 mutually exclusive ways each with probability $\frac{1}{2^4}$.

24. One child is selected at random from each of the three groups of children, namely, 3 girls and 1 boy, 2 girls and 2 boys, 1 girl and 3 boys. Find the probability of selecting 1 girl and 2 boys.

Hint: The event can occur in any one of the following mutually exclusive ways : BBG, BGB, GBB.

25. A can hit a target in 3 out of 4 attempts while B can hit it in 2 out of 3 attempts. If both of them try simultaneously, what is the probability that the target will be hit?

Hint: Find the probability of hitting the target at least once.

26. A and B played 12 chess matches out of which A won 6 matches, B won 4 matches and 2 resulted in draw. If they decide to play 3 more matches, what is the probability that (a) A wins all the three matches, (b) two matches end in draw, (c) B wins at least a match, (d) A wins at least a match, (e) A and B wins alternatively?

Hint: (b) $P(\text{two matches end in draw}) = \frac{2}{12} \times \frac{2}{12} \times \frac{10}{12} \times 3$.

27. A and B who are equally perfect players of badminton, stopped playing a match when their scores were 12 and 13 respectively. If 15 points are needed to win this match, what are their respective probabilities of winning?

Hint: A can win in following mutually exclusive ways; AAA, BAAA, ABAA, AABA.

28. A problem in accountancy is given to five students. Their chances of solving it are $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$ and $\frac{1}{6}$ respectively. What is the probability that the problem will be solved?

Hint: $P(A \cup B \cup C \cup D \cup E) = 1 - P(\bar{A}).P(\bar{B}).P(\bar{C}).P(\bar{D}).P(\bar{E})$.

29. (a) A guard of 12 soldiers is to be formed out of n soldiers. Find the probability that (i) two particular soldiers A and B are together on the guard, (ii) three particular soldiers C, D and E are together on the guard. (iii) Also find n if A and B are 3 times as often together on the guard as C, D and E.
- (b) A has 6 shares in a lottery in which there are 3 prizes and 10 blanks. B has 2 shares in a lottery in which there are 4 prizes and 8 blanks. Which of them has a better chance to win a prize?

Hint: (a) When A and B are on the guard, remaining 10 soldiers can be selected in ${}^{n-2}C_{10}$ ways.

$$(b) P(A) = 1 - \frac{{}^{10}C_6}{{}^{13}C_6}.$$

30. It is 8 to 5 against a person, who is now 40 years old, living till he is 70 and 4 to 3 against a person, now 50 years old, living till he is 80. Find the probability that at least one of them would be alive 30 years hence.

Hint: See example 37.

31. A candidate is selected for interview for 3 posts. There are 3 candidates for the first, 4 for the second and 2 for the third post. What are the chance of his getting at least one post?

Hint: Probability that he gets the first post is $\frac{1}{3}$, etc.

32. A bag contains 6 Rupee and 9 Dollar coins. Two drawings of 4 coins each are made without replacement. What is the probability that first draw will give 4 Rupee coins and second 4 dollar coins?

Hint: See example 46.

33. Three tokens marked as 1, 2 and 3 are placed in a bag and one is drawn and replaced. The operation being repeated three times. What is the probability of obtaining a total of 6?

Hint: A total of 6 can be obtained if different number is obtained in each operation or 2 is obtained in all the three operations. There are $3!$ ways of obtaining different numbers.

34. A certain player, say X, is known to win with probability 0.3 if the track is fast and with probability 0.4 if the track is slow. On Monday, there is a 0.7 probability of a fast track. What is the probability that X will win on Monday?

Hint: Let A be the event that the track is fast and B be the event that X wins, then

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

35. The probability that a vacuum-cleaner salesman will succeed in persuading a customer on the first call is 0.4. If he fails, the probability of success on the second call is 0.2. If he fails on the first two calls, the probability of success on the third and last call is 0.1. Find the probability that the salesman makes a sale of vacuum-cleaner to a customer.

Hint: Try as in exercise 34 above.

36. There are two contractors A and B, for the completion of a project. Contractor A does the first part of the project and then contractor B, by doing the second part, completes the project. B cannot start until A has finished. If A finishes on time, B has 85% chance of completing the project on time. If A doesn't finish on time, then

B has only 30% chance of completing the project on time. If A has 70% chance of finishing his work on time, what is the probability that the project will be finished on time?

Hint: Find $P(A \cap B) + P(\bar{A} \cap B)$.

37. The probability that a person stopping at a petrol pump will ask to have his tyres checked is 0.12, the probability that he will ask to have his oil checked is 0.29 and the probability that he will ask to have both of them checked is 0.07.
- What is the probability that a person stopping at the petrol pump will have either tyres or oil checked?
 - What is the probability that a person who has tyres checked will also have oil checked?
 - What is the probability that a person who has oil checked will also have tyres checked?

Hint: See example 32.

38. There are three brands, say X, Y and Z, of an item available in the market. A consumer chooses exactly one of them for his use. He never buys two or more brands simultaneously. The probabilities that he buys brands X, Y and Z are 0.20, 0.16 and 0.45 respectively.
- What is the probability that he doesn't buy any of the brands?
 - Given that the consumer buys some brand, what is the probability that he buys brand X?

Hint: (i) The required probability = $1 - P(X \cup Y \cup Z)$, where X, Y and Z are mutually exclusive.

39. A person applies for the post of manager in two firms A and B. He estimates that the probability of his being selected on firm A is 0.75, the probability of being rejected in firm B is 0.45 and the probability of rejection in at least one of the firms is 0.55. What is the probability that he will be selected in at least one of the firms?

Hint: $P(A \cap B) = 1 - P(\bar{A} \cup \bar{B})$.

40. (a) A student is given a true-false examination with 10 questions. If he gets 8 or more correct answers, he passes the examination. Given that he guesses at an answer to each question, compute the probability that he passes the examination.
- (b) In a multiple choice question, there are four alternative answers out of which one or more are correct. A candidate will get marks in the question only if he ticks all the correct answers. If he is allowed up to three chances to answer the question, find the probability that he will get marks in the question.

Hint: (a) $n(S) = 2^{10}$. No. of favourable cases is ${}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10}$.

- (b) Total no. of ways in which the student can tick the answers in one attempt = $2^4 - 1$ (since at least one of the answer is correct, therefore, it is not possible that he will leave all the answers unticked).

The total no. of ways of selecting three solutions from 15 is ${}^{15}C_3$. Note that it will be in the interest of the candidate to select a different solution in each attempt. Since out of 15 solutions, only one (way of marking the questions) is correct, therefore, the no. of ways of selecting incorrect solutions is ${}^{14}C_3$.

Hence the required probability is given by $1 - \frac{{}^{14}C_3}{{}^{15}C_3}$.

41. 200 students were admitted to an under graduate course through an entrance test out of which only 150 completed it successfully. On the examination of their admission data, it was found that 70% of those who passed and 50% of those who failed had a first division in their senior secondary examination. Find (a) the probability that a student with first division in the senior secondary examination is successful in the under graduate course, (b) the probability that a student without first division in senior secondary examination, is successful in the under graduate course, (c) the probability that an admitted student is a first divisioner in senior secondary examination, (d) the probability that an admitted student is unsuccessful in the under graduate course.

Hint: See example 27.

42. 300 employees of a firm were asked if they would favour increasing their working day by one hour so that they could have a five day week. The results are given in the following table :

	Favour(F)	Disfavour(D)	Neutral(N)
Men(M)	102	90	48
Women(W)	42	6	12

Find (a) $P(M)$, (b) $P(W)$, (c) $P(F)$, (d) $P(D)$, (e) $P(N)$, (f) $P(M \cap F)$, (g) $P(W \cap F)$, (h) $P(N \cap M)$, (i) $P(W \cap N)$, (j) $P(F/M)$, (k) $P(W/F)$, (l) $P(D/W)$, (m) $P(M/N)$, (n) $P(N/W)$, (o) $P(M \cap F)$, (p) $P(W \cup D)$, (q) $P(M \cup D)$, (r) $P(F \cup D)$, (s) $P(M \cup W)$, (t) $P(M \cup F \cup D)$.

Hint: See example 27.

43. In a bridge game of playing cards, 4 players are distributed one card each by turn so that each player gets 13 cards. What is the probability that a specified player gets a black ace and a king?

Hint: No. of favourable cases are ${}^2C_1 \times {}^4C_1 \times {}^{46}C_{11}$.

44. A bag contains 4 white and 2 black balls. Two balls are drawn successively one after another without replacement. What is the probability that (a) the first ball is white and the second is black, (b) the first is black and second is white.

Hint: Use conditional probability theorem.

45. (a) What is the probability that out of 3 friends, Ram, Shyam and Mohan, at least two have the same birthday?
 (b) What is the probability that out of a group of 4 persons, all born in the month of April, at least three have same birthday?

Hint: Suppose that Ram states his birthday, then the probability of Shyam having a

different birthday is $\frac{364}{365}$ and then the probability of Mohan having a different

birthday is $\frac{363}{365}$, etc. The required probability is $1 - \frac{364}{365} \times \frac{363}{365}$.

46. The probability that a man aged 70 years will die in a year is $\frac{2}{3}$. Find the probability that out of 5 men A_1, A_2, A_3, A_4 and A_5 , each aged 70 years, A_1 will die in a year and will be the first to die.

Hint: $P(A_1 \text{ dies first out of 5 men}) = \frac{1}{5}$. Multiply this by the probability that at least one of them die in a year.

47. The probability of rain tomorrow is 0.65 and the probability that the temperature will rise above 35°C is 0.8. The probability there is no rain and temperature remaining below 35°C is 0.1.

- (a) What is the probability of rain if temperature rises above 35°C ?
- (b) What is the probability that temperature remains below 35°C , given that there is no rain?

Hint: Try as in exercise 38 above.

48. A bag contains 4 red and 2 black balls. Three men X, Y and Z draw a ball in succession, without replacement, until a black ball is obtained. Find their respective chances of getting first black ball.

Hint: X can get first black ball in the following two mutually exclusive ways: B or WWWB, etc.

49. A and B are two candidates for admission to a certain course. The probability that A is selected is 0.80 and the probability that both A and B are selected is at the most 0.25. Is it possible that probability of selection of B is 0.50?

Hint: $P(A \cup B) \leq 1$.

50. Delhi has three independent reserved sources of electric power to use to prevent a blackout in the event that its regular source fails. The probability that any reserved source is available when its regular source fails is 0.7. What is the probability of not having a blackout if the regular source fails?

Hint: The required probability = 1 - the probability that power is not available from any of the reserved sources.

51. In a locality, out of 5,000 people residing, 1,200 are above 30 years of age and 3,000 are females. Out of 1,200, who are above 30 years, 200 are females. If a person selected at random is a female, what is the probability that she is above 30 years of age?

Hint: See example 27.

52. The probability that both the events A and B occur simultaneously is $\frac{1}{5}$ and the probability of occurrence of neither of them is $\frac{4}{15}$. Find the probabilities $P(A)$ and $P(B)$ on the assumption that the events are independent.

Hint: Let $P(A) = x$ and $P(B) = y$. Use the equation $1 - P(A \cup B) = P(\bar{A} \cap \bar{B})$ to find $x + y$. Find $x - y$ from it by using the equation $(x - y)^2 = (x + y)^2 - 4xy$.

53. Two factories A and B manufacture the same machine part. Each part is classified as having 0, 1, 2 or 3 manufacturing defects. The joint probabilities are as follows:

	Number of defects			
	0	1	2	3
Factory A	0.1250	0.0625	0.1875	0.1250
Factory B	0.0625	0.0625	0.1250	0.2500

- (i) A part is observed to have no defects. What is the probability that it was produced by factory A?
- (ii) A part is known to have been produced by factory A. What is the probability that the part has no defects?
- (iii) A part is known to have two or more defects. What is the probability that it was manufactured by factory A?

- (iv) A part is known to have one or more defects. What is the probability that it was manufactured by factory B?

Hint: See example 30.

54. A man is dealt 4 spade cards from an ordinary pack of 52 cards. If he is given three more cards, find the probability that at least one of the additional cards is also a spade.

Hint: The probability that no spade is obtained from the remaining 48 cards

$$\text{is } \frac{\binom{39}{3}}{\binom{48}{3}}.$$

55. An unbiased die is thrown three times. Find the probability of (a) throwing 4 on the first die if the sum of numbers obtained in three throws is 15, (b) obtaining a sum of 15 when first die shows 4.

Hint: (a) There are 10 ways of obtaining the sum 15 out of which 2 are favourable, (b) there are 36 cases in which first die shows 4, out of which only two are favourable.

56. A committee of 4 has to be formed from among 3 economists, 4 engineers, 2 statisticians and 1 doctor.

- (i) What is the probability that each of the four professions are represented on the committee?
- (ii) What is the probability that the committee consists of doctor and at least one economist?

Hint: (ii) The required probability is obtained by finding the probabilities of the following mutually exclusive events : {1 doc, 1 eco, 2 others}, {1 doc, 2 eco, 1 other} and {1 doc, 3 eco}.

57. Six persons toss a coin turn by turn. The game is won by the player who first throws a head. Find the probability of success of the fifth player.

Hint: See example 39.

58. Find the probability that an assessee files his tax return and cheats on it, given that 70% of all the assessee files returns and 20%, of those who file, cheat.

Hint: See example 27.

59. Two persons A and B throw three unbiased dice. If A throws 14, find B's chances of throwing a higher number.

Hint: The event that A throws 14 is independent of the event that B throws a higher number.

60. A is one of 6 horses entered for a race and is to be ridden by one of the jockeys B and C. It is 2 : 1 that B rides A, in which case all the horses are equally likely to win; if C rides A, his chances are trebled; what are the odds against his winning?

$$\text{Hint: } P(A \text{ wins given that he is ridden by jockey B}) = \frac{1}{6}$$

$$P(A \text{ wins given that he is ridden by jockey C}) = \frac{3}{6}$$

61. What is the probability that over a two day period the number of requests would either be 11 or 12 if at a motor garage the records of service requests alongwith their probabilities are given below?

Daily demand :	5	6	7
Probability :	0.25	0.65	0.10

Hint: 11 requests can occur in 2 ways and 12 requests in 3 ways.

62. The probability that T.V. of a company fails during first month of its use is 0.02. Of those that do not fail during first month, the probability of failure in the next five months is 0.01. Of those that do not fail during the first six months, the probability of failure by the end of the first year is 0.001. The company replaces, free of charge, any set that fails during its warranty period. If 2,000 sets are sold, how many will have to be replaced if the warranty period is (a) six months, (b) one year?

Probability

Hint: Probability that a set fails during first year = $0.02 + 0.98 \times 0.01 + 0.9902 \times 0.001$.

63. A salesman has 60% chances of making sales to each customer. The behaviour of each successive customer is assumed to be independent. If two customers A and B enter, what is the probability that the salesman will make sales to A or B?

Hint: $P(A \cup B) = 1 - P(\bar{A} \cap \bar{B})$.

64. A box contains 24 bulbs out of which 4 are defective. A customer draws a sample of 3 bulbs at random in succession and rejects the box if the sample contains one or more defectives. What is the probability that the box is rejected?

Hint: The box will be rejected if the sample contains at least one defective.

10.7 THEOREMS ON PROBABILITY - II

Theorem 5: (Bayes' Theorem or Inverse Probability Rule): The probabilities assigned to various events on the basis of the conditions of the experiment or by actual experimentation or past experience or on the basis of personal judgement are called *prior probabilities*. One may like to revise these probabilities in the light of certain additional or new information. This can be done with the help of Bayes' Theorem, which is based on the concept of conditional probability. The revised probabilities, thus obtained, are known as *posterior* or *inverse probabilities*. Using this theorem it is possible to revise various business decisions in the light of additional information.

Bayes' Theorem

If an event D can occur only in combination with any of the n mutually exclusive and exhaustive events A_1, A_2, \dots, A_n and if, in an actual observation, D is found to have occurred, then the probability that it was preceded by a particular event A_k is given by

$$P(A_k / D) = \frac{P(A_k) \cdot P(D / A_k)}{\sum_{i=1}^n P(A_i) \cdot P(D / A_i)}$$

Proof: Since A_1, A_2, \dots, A_n are n exhaustive events, therefore,

$$S = A_1 \cup A_2 \cup \dots \cup A_n.$$

Since D is another event that can occur in combination with any of the mutually exclusive and exhaustive events A_1, A_2, \dots, A_n , we can write

$$D = (A_1 \cap D) \cup (A_2 \cap D) \cup \dots \cup (A_n \cap D)$$

Taking probability of both sides, we get

$$P(D) = P(A_1 \cap D) + P(A_2 \cap D) + \dots + P(A_n \cap D)$$

We note that the events $(A_1 \cap D), (A_2 \cap D), \dots$ etc. are mutually exclusive.

$$P(D) = \sum_{i=1}^n P(A_i \cap D) = \sum_{i=1}^n P(A_i) \cdot P(D / A_i) \quad \dots (1)$$

The conditional probability of an event A_k given that D has already occurred, is given by

$$P(A_k / D) = \frac{P(A_k \cap D)}{P(D)} = \frac{P(A_k) \cdot P(D / A_k)}{P(D)} \quad \dots (2)$$

Substituting the value of $P(D)$ from (1), we get

$$P(A_k / D) = \frac{P(A_k) \cdot P(D / A_k)}{\sum_{i=1}^n P(A_i) \cdot P(D / A_i)} \quad \dots (3)$$

Example 54: A manufacturing firm purchases a certain component, for its manufacturing process, from three sub-contractors A, B and C. These supply 60%, 30% and 10% of the firm's requirements, respectively. It is known that 2%, 5% and 8% of the items supplied by the respective suppliers are defective. On a particular day, a normal shipment arrives from each of the three suppliers and the contents get mixed. A component is chosen at random from the day's shipment :

- (a) What is the probability that it is defective?
- (b) If this component is found to be defective, what is the probability that it was supplied by (i) A, (ii) B, (iii) C ?

Solution: Let A be the event that the item is supplied by A. Similarly, B and C denote the events that the item is supplied by B and C respectively. Further, let D be the event that the item is defective. It is given that :

$$P(A) = 0.6, P(B) = 0.3, P(C) = 0.1, P(D/A) = 0.02$$

$$P(D/B) = 0.05, P(D/C) = 0.08.$$

- (a) We have to find $P(D)$

From equation (1), we can write

$$P(D) = P(A \cap D) + P(B \cap D) + P(C \cap D)$$

$$= P(A)P(D/A) + P(B)P(D/B) + P(C)P(D/C)$$

$$= 0.6 \times 0.02 + 0.3 \times 0.05 + 0.1 \times 0.08 = 0.035$$

- (b) (i) We have to find $P(A/D)$

$$P(A/D) = \frac{P(A)P(D/A)}{P(D)} = \frac{0.6 \times 0.02}{0.035} = 0.343$$

$$\text{Similarly, (ii)} \quad P(B/D) = \frac{P(B)P(D/B)}{P(D)} = \frac{0.3 \times 0.05}{0.035} = 0.429$$

$$\text{and (iii)} \quad P(C/D) = \frac{P(C)P(D/C)}{P(D)} = \frac{0.1 \times 0.08}{0.035} = 0.228$$

Alternative Method: The above problem can also be attempted by writing various probabilities in the form of following table :

	A	B	C	Total
D	$P(A \cap D) = 0.012$	$P(B \cap D) = 0.015$	$P(C \cap D) = 0.008$	0.035
\bar{D}	$P(A \cap \bar{D}) = 0.588$	$P(B \cap \bar{D}) = 0.285$	$P(C \cap \bar{D}) = 0.092$	0.965
Total	0.600	0.300	0.100	1.000

Thus $P(A/D) = \frac{0.012}{0.035}$ etc.

Example 55: A box contains 4 identical dice out of which three are fair and the fourth is loaded in such a way that the face marked as 5 appears in 60% of the tosses. A die is selected at random from the box and tossed. If it shows 5, what is the probability that it was a loaded die?

Solution: Let A be the event that a fair die is selected and B be the event that the loaded die is selected from the box.

Then, we have $P(A) = \frac{3}{4}$ and $P(B) = \frac{1}{4}$.

Further, let D be the event that 5 is obtained on the die, then

$$P(D/A) = \frac{1}{6} \text{ and } P(D/B) = \frac{6}{10}$$

$$\text{Thus, } P(D) = P(A).P(D/A) + P(B).P(D/B) = \frac{3}{4} \times \frac{1}{6} + \frac{1}{4} \times \frac{6}{10} = \frac{11}{40}$$

We want to find $P(B/D)$, which is given by

$$P(B/D) = \frac{P(B \cap D)}{P(D)} = \frac{1}{4} \times \frac{6}{10} \times \frac{40}{11} = \frac{6}{11}$$

Example 56: A bag contains 6 red and 4 white balls. Another bag contains 3 red and 5 white balls. A fair die is tossed for the selection of bag. If the die shows 1 or 2, the first bag is selected otherwise the second bag is selected. A ball is drawn from the selected bag and is found to be red. What is the probability that the first bag was selected?

Solution: Let A be the event that first bag is selected, B be the event that second bag is selected and D be the event of drawing a red ball.

Then, we can write

$$P(A) = \frac{1}{3}, P(B) = \frac{2}{3}, P(D/A) = \frac{6}{10}, P(D/B) = \frac{3}{8}$$

$$\text{Further, } P(D) = \frac{1}{3} \times \frac{6}{10} + \frac{2}{3} \times \frac{3}{8} = \frac{9}{20}.$$

$$\therefore P(A/D) = \frac{P(A \cap D)}{P(D)} = \frac{1}{3} \times \frac{6}{10} \times \frac{20}{9} = \frac{4}{9}$$

Example 57: In a certain recruitment test there are multiple-choice questions. There are 4 possible answers to each question out of which only one is correct. An intelligent student knows 90% of the answers while a weak student knows only 20% of the answers.

(i) An intelligent student gets the correct answer, what is the probability that he was guessing?

(ii) A weak student gets the correct answer, what is the probability that he was guessing?

Solution: Let A be the event that an intelligent student knows the answer, B be the event that the weak student knows the answer and C be the event that the student gets a correct answer.

(i) We have to find $P(\bar{A}/C)$. We can write

$$P(\bar{A}/C) = \frac{P(\bar{A} \cap C)}{P(C)} = \frac{P(\bar{A})P(C/\bar{A})}{P(\bar{A})P(C/\bar{A}) + P(A)P(C/A)} \quad \dots (1)$$

It is given that $P(A) = 0.90$, $P(C/A) = \frac{1}{4} = 0.25$ and $P(C/\bar{A}) = 1.0$

From the above, we can also write $P(\bar{A}) = 0.10$

Substituting these values, we get

$$P(\bar{A}/C) = \frac{0.10 \times 0.25}{0.10 \times 0.25 + 0.90 \times 1.0} = \frac{0.025}{0.925} = 0.027$$

- (ii) We have to find $P(\bar{B}/C)$. Replacing \bar{A} by \bar{B} , in equation (1), we can get this probability.

It is given that $P(B) = 0.20$, $P(C/B) = 0.25$ and $P(C/\bar{B}) = 1.0$

From the above, we can also write $P(\bar{B}) = 0.80$

$$\text{Thus, we get } P(\bar{B}/C) = \frac{0.80 \times 0.25}{0.80 \times 0.25 + 0.20 \times 1.0} = \frac{0.20}{0.40} = 0.50$$

Example 58: An electronic manufacturer has two lines A and B assembling identical electronic units. 5% of the units assembled on line A and 10% of those assembled on line B are defective. All defective units must be reworked at a significant increase in cost. During the last eight-hour shift, line A produced 200 units while the line B produced 300 units. One unit is selected at random from the 500 units produced and is found to be defective. What is the probability that it was assembled (i) on line A, (ii) on line B?

Answer the above questions if the selected unit was found to be non-defective.

Solution: Let A be the event that the unit is assembled on line A, B be the event that it is assembled on line B and D be the event that it is defective.

Thus, we can write

$$P(A) = \frac{2}{5}, P(B) = \frac{3}{5}, P(D/A) = \frac{5}{100} \text{ and } P(D/B) = \frac{10}{100}$$

Further, we have

$$P(A \cap D) = \frac{2}{5} \times \frac{5}{100} = \frac{1}{50} \text{ and } P(B \cap D) = \frac{3}{5} \times \frac{10}{100} = \frac{3}{50}$$

The required probabilities are computed form the following table:

	A	B	Total
D	$\frac{1}{50}$	$\frac{3}{50}$	$\frac{4}{50}$
\bar{D}	$\frac{19}{50}$	$\frac{27}{50}$	$\frac{46}{50}$
Total	$\frac{20}{50}$	$\frac{30}{50}$	1

From the above table, we can write

$$P(A/D) = \frac{1}{50} \times \frac{50}{4} = \frac{1}{4}, P(B/D) = \frac{3}{50} \times \frac{50}{4} = \frac{3}{4}$$

$$P(A/\bar{D}) = \frac{19}{50} \times \frac{50}{46} = \frac{19}{46}, P(B/\bar{D}) = \frac{27}{50} \times \frac{50}{46} = \frac{27}{46}$$

Exercise with Hints:

- An insurance company insured 2,000 scooter drivers, 4,000 car drivers and 6,000 truck drivers. The probability of an accident is 0.01, 0.03 and 0.15 in the respective category. One of the insured driver meets an accident. What is the probability that he is a scooter driver?

Hint: Apply Bayes' Rule.

- When a machine is set correctly, it produces 25% defectives, otherwise it produces 60% defectives. From the past knowledge and experience, the manufacturer knows that the chance that the machine is set correctly or wrongly is 50 : 50. The machine was set before the commencement of production and 1 piece was taken out and found to be defective. What is the probability of the machine set up being correct? If the selected piece was found to be non-defective, what is the probability of the machine set up being wrong?

Hint: Apply Bayes' Rule.

- Each of the three identical jewellery boxes has 2 drawers. In each drawer of the first box there is a gold watch. In each drawer of the second box there is a silver watch. In one drawer of the third box there is a gold watch while in the other drawer there is a silver watch. If we select a box at random, open one of the drawers and find it to contain a silver watch, what is the probability that the other drawer has a gold watch?

$$\text{Hint: } P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}, \quad P(S/B_1) = 0, \quad P(S/B_2) = 1, \quad P(S/B_3) = \frac{1}{2}.$$

- In a factory producing bolts, Machines A, B and C manufacture 25%, 35% and 40% of total output. Of their output, 5%, 4% and 2% are defective respectively. A bolt is drawn at random from the product and is found to be defective. What is the probability that it was manufactured by machine A?

Hint: Apply Bayes' Rule.

- Consider a population of consumers consisting of two types. The upper class of consumers comprise 35% of the population and each member has a probability 0.8 of purchasing brand A of a product. Each member of the rest of the population has a probability 0.3 of purchasing brand A of the product. A consumer, chosen at random, is found to be buyer of brand A. What is the probability that the buyer belongs to the middle and lower class of consumers?

Hint: Apply Bayes' Rule.

- At an electric plant, it is known from the past experience that the probability is 0.86 that new worker who has attended the company's training programme will meet his production quota and that the corresponding probability is 0.35 for a new worker who has not attended the company's training programme. If 80% of the new workers attend the training programme, what is the probability that new worker will meet his production quota?

$$\text{Hint: Apply } P(D) = P(A).P(D/A) + P(B).P(D/B)$$

- A talcum powder manufacturing company had launched a new type of advertisement. The company estimated that a person who comes across the advertisement will buy their product with a probability of 0.7 and those who does not see the advertisement will buy the product with a probability of 0.3. If in an area of 1,000 people, 70% had come across the advertisement, what is the probability that a person who buys the product (a) has not come across the advertisement (b) has come across the advertisement?

Hint: Apply Bayes' Rule.

8. There are two boxes, of identical appearance, each containing 4 sparkplugs. It is known that box I contains only one defective sparkplug, while all the four sparkplugs of box II are non-defective. A sparkplug is drawn at random from a box, selected at random, is found to be non-defective. What is the probability that it came from box I?

Hint: Apply Bayes' Rule.

9. A man has 5 one rupee coins and one of them is known to have two heads. He takes out a coin at random and tosses it 5 times; it always falls head upward. What is the probability that it is a coin with two heads?

Hint: Apply Bayes' Rule.

Check Your Progress 10.2

- 1 Give Statistical or Empirical Definition of Probability?
2. Explain Permutation with restrictions.

Notes: (a) Write your answer in the space given below.
(b) Please go through the lesson sub-head thoroughly you will get your answers in it.
(c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

10.8 LET US SUM UP

Probability distributions are a fundamental concept in statistics. They are used both on a theoretical level and a practical level. Formally, a probability is a bundle of four things: a universe which is a set of all possible results, a number which is an extension of a Boolean truth value, a constraint which matches the logical law of the excluded middle along with the few operations an arithmetical operations correspond to the standard logical operation of Boolean logic.

At last we can say that a probability is a measure of the likelihood of an event. It can be elucidate as a decimal or a percentage. Every probability must be between 0 and 1 (100%) inclusive P=0 indicate an impossible event, P=1 or 100% indicates a certainty

1. (a) The number of permutations of n objects taking n at a time are n!
- (b) The number of permutations of n objects taking r at a time, are ${}^n P_r = \frac{n!}{(n-r)!}$
- (c) The number of permutations of n objects in a circular order are $(n - 1)!$
- (d) The number of permutations of n objects out of which n_1 are alike, n_2 are alike, n_k are alike, are $\frac{n!}{n_1! n_2! \dots n_k!}$

- (e) The number of combinations of n objects taking r at a time are Probability
- $${}^n C_r = \frac{n!}{r!(n-r)!}$$
2. (a) The probability of occurrence of at least one of the two events A and B is given by : $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1 - P(\bar{A} \cap \bar{B})$.
- (b) The probability of occurrence of exactly one of the events A or B is given by: $P(A \cap \bar{B}) + P(\bar{A} \cap B)$ or $P(A \cup B) - P(A \cap B)$
3. (a) The probability of simultaneous occurrence of the two events A and B is given by: $P(A \cap B) = P(A).P(B/A)$ or $= P(B).P(A/B)$
- (b) If A and B are independent $P(A \cap B) = P(A).P(B)$.
4. Bayes' Theorem :

$$P(A_k / D) = \frac{P(A_k).P(D/A_k)}{\sum_{i=1}^n P(A_i).P(D/A_i)}, \quad (k = 1, 2, \dots, n)$$

Here A_1, A_2, \dots, A_n are n mutually exclusive and exhaustive events.

10.9 LESSON-END ACTIVITY

Apply the concept of probability in predicting the sensex in different stock exchanges like National Stock Exchange, Delhi Stock Exchange and Bombay Stock Exchange.

10.10 KEYWORDS

Probability

Event

Outcome

Occurrence

Combination

Inverse probability

10.11 QUESTIONS FOR DISCUSSION

1. Fill in the blank

- (a) The theory of probability is a study of or experiments.
- (b) is a arrangement of a given number of objects in a definite order.
- (c) Counting techniques are often helpful in of total no. of outcomes.
- (d) Modern approach was introduced by mathematical.
- (e) Bayes' Theorem is also called

2. Distinguish between

- (a) Favourable outcomes and Exhaustive outcomes

- (b) Permutation and Combination
 - (c) Prior possibilities or Inverse possibilities
3. **Write True or False against each of the statement:**
- (a) It is not possible to predetermine the outcome association with a particular experimentation.
 - (b) The total no. of permutations of n distinct objects is $n!$
 - (c) Each element of the set is called sample point.
 - (d) A compound event is simultaneous occurrence of only two events.
 - (e) The assignment of probabilities on basis of statistical and classical events is objective.

10.12 TERMINAL QUESTIONS

1. Define the term 'probability' by (a) The Classical Approach, (b) The Statistical Approach. What are the main limitations of these approaches?
2. Discuss the axiomatic approach to probability. In what way it is an improvement over classical and statistical approaches?
3. Distinguish between objective probability and subjective probability. Give one example of each concept.
4. State and prove theorem of addition of probabilities for two events when (a) they are not independent, (b) they are independent.
5. Explain the meaning of conditional probability. State and prove the multiplication rule of probability of two events when (a) they are not independent, (b) they are independent.
6. Explain the concept of independence and mutually exclusiveness of two events A and B. If A and B are independent events, then prove that \bar{A} and \bar{B} are also independent.

For two events A and B it is given that $P(A) = 0.4$, $P(B) = p$, $P(A \cup B) = 0.6$

- (i) Find the value of p so that A and B are independent.
 - (ii) Find the value of p so that A and B are mutually exclusive.
7. Explain the meaning of a statistical experiment and corresponding sample space. Write down the sample space of an experiment of simultaneous toss of two coins and a die.
 8. State and prove Bayes' theorem on inverse probability.
 9. What is the probability of getting exactly two heads in three throws of an unbiased coin?
 10. What is the probability of getting a sum of 2 or 8 or 12 in single throw of two unbiased dice?
 11. Two cards are drawn at random from a pack of 52 cards. What is the probability that the first is a king and second is a queen?
 12. What is the probability of successive drawing of an ace, a king, a queen and a jack from a pack of 52 well shuffled cards? The drawn cards are not replaced.
 13. 5 unbiased coins with faces marked as 2 and 3 are tossed. Find the probability of getting a sum of 12.

14. If 15 chocolates are distributed at random among 5 children, what is the probability that a particular child receives 8 chocolates?
15. A and B stand in a ring with 10 other persons. If arrangement of 12 persons is at random, find the chance that there are exactly three persons between A and B.
16. Two different digits are chosen at random from the set 1, 2, 3, 4, 5, 6, 7, 8. Find the probability that sum of two digits exceeds 13.
17. From each of the four married couples one of the partner is selected at random. What is the probability that they are of the same sex?
18. A bag contains 5 red and 4 green balls. Two draws of three balls each are done with replacement of balls in the first draw. Find the probability that all the three balls are red in the first draw and green in the second draw.
19. Two die are thrown two times. What is the probability of getting a sum 10 in the first and 11 in the second throw?
20. 4 cards are drawn successively one after the other without replacement. What is the probability of getting cards of the same denominations?
21. A bag contains 4 white and 2 black balls. Two balls are drawn one after another without replacement. What is the probability that first ball is white and second is black or first is black and second is white?
22. A bag contains 4 white and 3 red balls. Another bag contains 3 white and 5 red balls. One ball is drawn at random from each bag. What is the probability that (a) both balls are white, (b) both are red, (c) one of them is white and the other is red?
23. What is the probability of a player getting all the four aces, when playing cards are uniformly distributed among the four players?
24. A bag contains 10 white and 6 red balls. Two balls are drawn one after another with replacement. Find the probability that both balls are red.
25. Three persons A, B and C successively draw one card from a pack of 52 cards with replacement of the card drawn earlier. The first to obtain a card of spade wins. What are their respective chances of winning?
26. A bag contains 6 red and 4 green balls. A ball is drawn at random and replaced and a second ball is drawn at random. Find the probability that the two balls drawn are of different colours.
27. The letters of the word GANESH PURI are arranged at random. Find the probability that in the word, so formed;
 - (a) The letter G always occupies the first place.
 - (b) The letter P and I respectively occupy first and last places.
 - (c) The vowels are always together.
 - (d) The letters E, H, P are never together.
 - (e) The vowels always occupy even places (i.e., 2nd, 4th, etc.)
28. 5-letter words are formed from the letters of the word ORDINATES. What is the probability that the word so formed consists of 2 vowels and 3 consonants?
29. Maximum number of different committees are formed out of 100 teachers, including principal, of a college such that each committee consists of the same number of members. What is the probability that principal is a member of any committee?
30. Letters of the word INTERMEDIATE are arranged at random to form different words. What is the probability that :

- (a) First letter of the word is R?
- (b) First letter is M and last letter is E?
- (c) All the vowels come together?
- (d) The vowels are never together?

Hint: (d) The event will occur if the letters are arranged as VCVCVCVCVCVCV where V and C denote vowels and consonants respectively. 6 places for vowels can be chosen in 7C_6 .

31. Five persons entered the lift cabin on the ground floor of an 8-floor building. Suppose that each of them independently and with equal probability can leave the floor beginning with first. Find out the probability of all the persons leaving at different floors.

Hint: There are 7 floors along with ground floor.

- 32. A team of first eleven players is to be selected at random from a group of 15 players. What is the probability that (a) a particular player is included, (b) a particular player is excluded?
- 33. Out of 18 players of a cricket club there are 2 wicket keepers, 5 bowlers and rest batsmen. What is the probability of selection of a team of 11 players including one wicket keeper and at least 3 bowlers?
- 34. Four persons are selected at random from a group consisting of 3 men, 2 women and 4 children. Find the chance that exactly 2 of them are children.
- 35. A committee of 6 is chosen from 10 men and 7 women so as to contain at least 3 men and 2 women. Find the probability that 2 particular women don't serve on the same committee.
- 36. If n persons are seated around a round table, find the probability that in no two ways a man has the same neighbours.
- 37. 6 teachers, of whom 2 are from science, 2 from arts and 2 from commerce, are seated in a row. What is the probability that the teachers of the same discipline are sitting together?
- 38. (a) If $P(A) = 0.5$, $P(B) = 0.4$ and $P(\bar{A} \cup B) = 0.7$, find $P(A/B)$ and $P(A \cup B)$, where \bar{A} is compliment of A. State whether A and B are independent.
 (b) If $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{2}$, $P(A/B) = \frac{1}{6}$, find $P(B/A)$ and $P(B/\bar{A})$.
 (c) If A, B and C are three mutually exclusive events, find $P(B)$ if

$$\frac{1}{3}P(C) = \frac{1}{2}P(A) = P(B).$$
- 39. Let A be the event that a business executive selected at random has stomach ulcer and B be the event that he has a heart disease. Interpret the following events :
 (i) $A^c \cup A$, (ii) $A^c \cap A$, (iii) $A^c \cap B$, (iv) $A \cap B^c$, (v) $(A \cap B)^c$,
 where c stands for compliment.
- 40. Let A, B and C be three events. Write down the following events in usual set notations :
 (i) A and B occur together, (ii) Both A and B occur but not C, (iii) all the three events occur, (iv) at least one event occur and (v) at least two events occur.

41. The records of 400 examinees are given below :

Probability

Score	Educational Qualification			Total
	B. A.	B. Sc.	B. Com.	
Below 50	90	30	60	180
Between 50 and 60	20	70	70	160
Above 60	10	30	20	60
Total	120	130	150	400

If an examinee is selected at random from this group, find

- (i) the probability that he is a commerce graduate,
 - (ii) the probability that he is a science graduate, given that his score is above 60 and
 - (iii) the probability that his score is below 50, given that he is B.A.
42. It is given that $P(A+B)=\frac{5}{6}$, $P(AB)=\frac{1}{3}$ and $P(\bar{B})=\frac{1}{2}$, where $P(\bar{B})$ stands for the probability that event B doesn't happen. Determine $P(A)$ and $P(B)$. Hence, show that the events A and B are independent.
43. A can solve 75% of the problems in accountancy while B can solve 70% of the problems. Find the probability that a problem selected at random from an accountancy book;
- (a) will be solved by both A and B,
 - (b) will be solved by A or B,
 - (c) will be solved by one of them.
44. (a) One card is drawn from each of two ordinary sets of 52 cards. Find the probability that at least one of them will be the ace of hearts.
- (b) Two cards are drawn simultaneously from a set of 52 cards. Find the probability that at least one of them will be the ace of hearts.
45. An article manufactured by a company consists of two parts X and Y. In the process of manufacture of part X, 9 out of 104 parts may be defective. Similarly, 5 out of 100 are likely to be defective in the manufacture of part Y. Compute the probability that the assembled product will not be defective.
46. A salesman has 80% chance of making a sale to each customer. The behaviour of each customer is independent. If two customers A and enter, what is the probability that the salesman will make a sale to A or B?
47. A problem in economics is given to 3 students whose chances of solving it are $\frac{2}{3}$, $\frac{3}{4}$ and $\frac{4}{5}$ respectively. What is the probability that the problem will be solved?
48. A man and a woman appear in an interview for two vacancies in the same post. The probability of man's selection is $\frac{1}{4}$ and that of woman's selection is $\frac{1}{3}$. What is the probability that
- (a) both of them will be selected?
 - (b) only one of them will be selected?
 - (c) none of them will be selected?
 - (d) at least one of them will be selected?

49. What is the chance that a non-leap year selected at random will contain 53 Sundays?
50. In a group of equal number of men and women 15% of men and 30% of women are unemployed. What is the probability that a person selected at random is employed?
51. An anti-aircraft gun can take a maximum of four shots at enemy plane moving away from it. The probabilities of hitting the plane at first, second, third and fourth shot are 0.4, 0.3, 0.2 and 0.1 respectively. What is the probability that the gun hits the plane?
52. A piece of equipment will function only when all the three components A, B, C are working. The probability of A failing during one year is 0.15 and that of B failing is 0.05 and of C failing is 0.10. What is the probability that the equipment will fail before the end of the year?
53. A worker attends three machines each of which operates independently of the other two. The probabilities of events that machines will not require operator's intervention during a shift are $p_1 = 0.4$, $p_2 = 0.3$ and $p_3 = 0.2$. Find the probability that at least one machine will require worker's intervention during a shift.
54. The probability that a contractor will get a plumbing contract is $2/3$ and the probability that he will not get a electric contract is $5/9$. If the probability of getting at least one of the contract is $4/5$, what is the probability that he will get both?
55. An M.B.A. applies for job in two firms X and Y. The probability of his being selected in firm X is 0.7 and being rejected in the firm Y is 0.5. The probability of at least one of his application being rejected is 0.6. What is the probability that he will be selected in one of the firms?
56. A researcher has to consult a recently published book. The probability of its being available is 0.5 for library A and 0.7 for library B. Assuming the two events to be statistically independent, find the probability of book being available in library A and not available in library B.
57. An investment consultant predicts that the odds against the price of certain stock will go up next week are 2:1 and odds in favour of price remaining same are 1:3. What is the probability that price of the stock will go down during the week?
58. In a random sample of 1,000 residents of a city 700 read newspaper A and 400 read newspaper B. If the habit of reading newspaper A and B is independent, what is the probability that a person selected at random would be reading (a) both the papers, (b) exactly one of the papers, (c) at least one of the papers? Also find the absolute number of persons in each of the cases (a), (b) and (c).
59. The odds that a book will be reviewed favourably by three independent experts are 5 to 2, 4 to 3 and 3 to 4 respectively. What is the probability that of the three reviews a majority will be favourable?
60. In a certain city two newspapers, A and B, are published. It is known that 25% of the city population reads A and 20% reads B while 8% reads both A and B. It is also known that 30% of those who read A but not B look into advertisements and 40% of those who read B but not A look into advertisements while 50% of those who read both A and B look into advertisements. What is percentage of population who reads an advertisement?
61. The probability that a new entrant to a college will be a student of economics is $1/3$, that he will be a student of political science is $7/10$ and that he will not be a student of economics and political science is $1/5$. If one of the new entrants is selected at random, what is the probability that (a) he will be a student of economics and political science, (b) he will be a student of economics if he is a student of

- political science? Comment upon the independence of two events : a student of economics and a student of political science.
62. 20% of all students at a university are graduates and 80% are undergraduates. The probability that a graduate student is married is 0.5 and the probability that an undergraduate student is married is 0.1. One student is selected at random.
 - (a) What is probability that he is married?
 - (b) What is the probability that he is a graduate if he is found to be married?
 63. In a city three daily news papers X, Y and Z are published. 40% of the people of the city read X, 50% read Y, 30% read Z, 20% read both X and Y, 15% read X and Z, 10% read Y and Z and 24% read all the 3 papers. Calculate the percentage of people who do not read any of the 3 newspapers.
 64. A bag contains 4 red and 3 blue balls. Two drawings of 2 balls are made. Find the probability of drawing first 2 red balls and the second 2 blue balls
 - (i) if the balls are returned to the bag after the first draw,
 - (ii) if the balls are not returned after the first draw.
 65. A die is loaded in such a way that each odd number is twice as likely to occur as each even number. Find (i) the probability that the number rolled is a perfect square and (ii) the probability that the number rolled is a perfect square provided it is greater than 3.
 66. There are 100 students in a college class of which 36 boys are studying statistics and 13 girls are not studying statistics. If there are 55 girls in all, find the probability that a boy picked at random is not studying statistics.
 67. If a pair of dice is thrown, find the probability that
 - (i) the sum is neither 7 nor 11
 - (ii) the sum is neither 8 nor 10
 - (iii) the sum is greater than 12.
 68. Three horses A, B and C are in race. A is twice as likely to win as B, and B is twice as likely to win as C. What are the respective probabilities of winning?
 69. A sample of 3 items is selected at random from a box containing 12 items of which 3 are defective. Find the possible number of defective combinations of the said 3 selected items along with their respective probabilities.
 70. In an examination 30% of students have failed in mathematics, 20% of the students have failed in chemistry and 10% have failed in both mathematics and chemistry. A student is selected at random.
 - (i) What is the probability that the student has failed either in mathematics or in chemistry?
 - (ii) What is the probability that the student has failed in mathematics if is known that he has failed in chemistry?
 71. There are two bags. The first contains 2 red and 1 white balls whereas the second bag contains 1 red and 2 white balls. One ball is taken out at random from the first bag and is being put in the second. Then, a ball is chosen at random from the second bag. What is the probability that this ball is red?
 72. From the sale force of 150 people, one will be chosen to attend a special meeting. If 52 are single and 72 are college graduates, and $\frac{3}{4}$ of 52 that are single are college graduates, what is the probability that a sales person, selected at random, will be neither single nor a college graduate?

73. Data on readership of a certain magazine indicate that the proportion of male readers over 30 years old is 0.20. The proportion of male readers under 30 is 0.40. If the proportion of readers under 30 is 0.70, what is the proportion of subscribers that are male? Also find the probability that a randomly selected male subscriber is under 30.
74. Two union leaders and 10 directors of a company sit randomly to decide upon the wage hike as demanded by the union. Find the probability that there will be exactly three directors between the two union leaders.
75. Suppose a company hires both MBAs and non-MBAs for the same kind of managerial task. After a period of employment some of each category are promoted and some are not. Table below gives the proportion of company's managers among the said classes :

Promotional Status	Academic Qualification		Total
	MBA	Non - MBA	
	(A)	(\bar{A})	
Promoted (B)	0.42	0.18	0.60
Not Promoted(\bar{B})	0.28	0.12	0.40
Total	0.70	0.30	1.00

- Calculate $P(A/B)$ and $P(B/A)$, and find out whether A and B are independent events?
76. Each of A, B and C throws with two dice for a prize. The highest throw wins, but if equal highest throws occur the player with these throw continue. If A throws 10 find his chance of winning.
 77. The probability of a man hitting a target is $1/4$. How many times must he fire so that probability of hitting the target at least once is greater than $2/3$?
 78. Find the probability that an assessee files his tax return and cheats on it, given that 70% of all assessee file returns and 25% of those who file, cheat.
 79. The probability of an aircraft engine failure is 0.10. With how many engines should the aircraft be equipped to be 0.999 sure against an engine failure? Assume that only one engine is needed for successful operation of the aircraft.
 80. A market research firm is interested in surveying certain attitude in small community. There are 125 house holds broken down according to income, ownership of a telephone and ownership of a T.V.

<i>Households with annual income of Rs. 1,00,000 or less</i>		<i>Households with annual income above Rs. 1,00,000</i>	
Telephone subscriber	No Telephone	Telephone subscriber	No Telephone
Own T.V. set	59	10	40
No T.V. set	2	4	1

- (a) If a person is selected at random, what is the probability that he is a T.V. owner?
- (b) If the person selected at random is found to be having income greater than 100,000 and a telephone subscriber, what is the probability that he is a T.V. owner?
- (c) What is the conditional probability of drawing a household that owns a T.V., given that he is a telephone subscriber?

81. An investment firm purchases three stocks for one week trading purposes. It assesses the probability that the stock will increase in value over the week as 0.8, 0.7 and 0.6 respectively. What is the chance that (a) all the three stocks will increase, and (b) at least two stocks will increase? (Assume that the movements of these stocks is independent.)
82. A company has two plants to manufacture scooters. Plant I manufactures 70% of the scooters and plant II manufactures 30%. At plant I 80% of the scooters are rated standard quality and at plant II 90% of the scooters are rated standard quality. A scooter is picked up at random and is found to be standard quality. What is the chance that it has been produced by plant I?
83. A person has 4 coins each of a different denomination. How many different sums of money can be formed?
84. Two sets of candidate avoid touching for the position of Board of Directors of a company. The probabilities of winning are 0.7 and 0.3 for the two. If the first set wins, they will introduce a new product with the probability 0.4. Similarly, the probability that the second set will introduce a new product is 0.8. If the new product has been introduced, what is the chance that the first set of candidates has won?
85. By examining the chest X-ray, the probability that T.B. is detected when a person is actually suffering is 0.99. The probability that the doctor diagnoses incorrectly, that a person has T.B., on the basis of X-ray is 0.001. In a certain city, 1 in 1000 persons suffers from T.B. A person selected at random is diagnosed to have T.B. What is the chance that he actually has T.B.?
86. The compressors used in refrigerators are manufactured by XYZ company at three factories located at Pune, Nasik and Nagpur. It is known that the Pune factory produces twice as many compressors as Nasik one, which produces the same number as the Nagpur one (during the same period). Experience also shows that 0.2% of the compressors produced at Pune and Nasik and 0.4% of those produced at Nagpur are defective.
- A quality control engineer while maintaining a refrigerator finds a defective compressor. What is the probability that Nasik factory is not to be blamed?
87. A company estimates that the probability of a person buying its product after seeing the advertisement is 0.7. If 60% of the persons have come across the advertisement, What is the probability that the person, who buys the product, has not come across the advertisement?
88. In an automobile factory, certain parts are to be fixed to the chassis in a section before it moves into another section. On a given day, one of the three persons A, B or C carries out this task. A has 45%, B has 35% and C has 20% chance of doing it. The probabilities that A, B or C will take more than the allotted time are $\frac{1}{16}$, $\frac{1}{10}$ and $\frac{1}{20}$ respectively. If it is found that one of them has taken more time, what is the probability that A has taken more time?
89. The probabilities of X, Y and Z becoming managers are $\frac{4}{9}$, $\frac{2}{9}$ and $\frac{1}{3}$ respectively. The probabilities that the Bonus Scheme will be introduced if X, Y or Z become manager are $\frac{3}{10}$, $\frac{1}{2}$ and $\frac{4}{5}$ respectively.

- (a) What is the probability that the Bonus Scheme will be introduced?
- (b) What is the probability that X was appointed as manager given that the Bonus Scheme has been introduced?

90. There are 3 bags. The first bag contains 5 red and 3 black balls, the second contains 4 red and 5 black balls and the third contains 3 red and 4 black balls. A bag is selected at random and the two balls drawn, at random, are found to be red. Revise the probabilities of selection of each bag in the light of this observation.
91. On an average, 20% of the persons going to a handicraft emporium are foreigners and the remaining 80% are local persons. 75% of foreigners and 50% of local persons are found to make purchases. If a bundle of purchased items is sent to the cash counter, what is the probability that the purchaser is a foreigner?
92. The chance that doctor A will diagnose disease B correctly is 60%. The chance that a patient will die by his treatment after correct diagnosis is 40% and the chance of death by wrong diagnosis is 70%. A patient of doctor A, who had disease B, died. What is the chance that his disease was correctly diagnosed?
93. A company has four production sections S_1 , S_2 , S_3 and S_4 which contribute 30%, 20%, 28% and 22%, respectively, to the total output. It was observed that these sections produced 1%, 2%, 3% and 4% defective units respectively. If a unit is selected at random and found to be defective, what is the probability that it has come from either S_1 or S_4 ?
94. A factory produces certain type of output by three machines. The respective daily production figures are : machine A = 3,000 units, machine B = 2,500 units, machine C = 4,500 units. Past experience shows that 1 % of the output produced by machine A is defective. The corresponding fractions of defectives for the other two machines are 1.2 and 2% respectively. An item is selected at random from a day's production and is found to be defective. What is the probability that it came from the output of (i) machine A, (ii) machine B, (iii) machine C?
95. It is known that 20% of the males and 5% of the females are unemployed in a certain town consisting of an equal number of males and females. A person selected at random is found to be unemployed. What is the probability that he/she is a (i) male, (ii) female?
96. In a typing-pool, three typists share the total work in the ratio 30%, 35% and 35% of the total work. The first, second and the third typist spoil the work to the extent of 3%, 4% and 5% respectively. A completed work is selected at random and found to be spoiled. What is the probability that the work was done by the third typist?
97. An organisation dealing with consumer products, wants to introduce a new product in the market. Based on their past experience, it has a chance of 65% of being successful and 35% of not being successful. In order to help them to make a decision on the new product, i.e., whether to introduce the new product or not, it decides to get additional information on consumers' attitude towards the product. For this purpose, the organisation decides to conduct a survey. In the past, when the product of this type were successful, the surveys yielded favourable indications 85% of the times, whereas unsuccessful products received favourable indications 30% of the time. Determine the probability of the product being a success given the survey information.
98. In a class of 75 students, 15 were considered to be very intelligent, 45 as medium and the rest below average. The probability that a very intelligent student fails in a viva-voce examination is 0.005; the medium student failing has a probability 0.05; and the corresponding probability for a below average student is 0.15. If a student is known to have passed the viva-voce examination, what is the probability that he is below average?

99. Comment on the following statements :

Probability

- (a) Since accident statistics show that the probability that a person will be involved in a road accident is 0.02, the probability that he will be involved in 2 accidents in that year is 0.0004.
- (b) For three mutually exclusive events A, B and C of a sample space S, where

$$P(A) = \frac{1}{3}, \quad P(B) = \frac{3}{5} \text{ and } P(C) = \frac{1}{5}.$$

- (c) A and B are two events in a sample space S where $P(A) = \frac{5}{6}$, $P(B) = \frac{2}{3}$ and $P(A \cap B) = \frac{2}{5}$.

- (d) Four persons are asked the same question by an interviewer. If each has, independently, probability of 1/6 of answering correctly, the probability that at least one answers correctly is $4 \times \frac{1}{6} = \frac{2}{3}$.

- (e) The probability that A and B, working independently, will solve a problem is $\frac{2}{3}$ and probability that A will solve the problem $\frac{1}{3}$.
- (f) For a biased dice the probabilities for different faces to turn up are as given in the following table:

Number on the dice	1	2	3	4	5	6
Probability	0.15	0.30	0.17	0.25	0.08	0.07

- (g) If the probability of A to fail in an examination is 0.15 and that for B is 0.27, then the probability that either A or B fails in examination is 0.42.
- (h) If the probability that Congress wins from a constituency is 0.40 and that B.J.P. wins from the same constituency is 0.42, than the probability that either Congress or B.J.P. wins from that constituency is 0.82.
- (i) The probability of occurrence of event A is 0.6 and the probability of occurrence of at least one of the four events A, B, C and D is 0.5.

100. Four alternative answers are given to each question. Point put the correct answer :

- (a) If A and B are any two events of a sample space S, then $P(A \cup B) + P(A \cap B)$ equals
 - (i) $P(A) + P(B)$
 - (ii) $1 - P(\bar{A} \cap \bar{B})$
 - (iii) $1 - P(\bar{A} \cup \bar{B})$
 - (iv) none of the above.
- (b) If A and B are independent and mutually exclusive events, then
 - (i) $P(A) = P(A/B)$
 - (ii) $P(B) = P(B/A)$

- (iii) either $P(A)$ or $P(B)$ or both must be zero.
 - (iv) none of the above.
- (c) If A and B are independent events, then $P(A \cap B)$ equals
- (i) $P(A) + P(B)$
 - (ii) $P(A) \cdot P(B/A)$
 - (iii) $P(B) \cdot P(A/B)$
 - (iv) $P(A) \cdot P(B)$
- (d) If A and B are independent events, then $P(A \cup B)$ equals
- (i) $P(A) \cdot P(B) - P(B)$
 - (ii) $P(A) \cdot P(\bar{B}) - P(B)$
 - (iii) $P(\bar{A}) \cdot P(\bar{B}) - P(A)$
 - (iv) none of the above.
- (e) If A and B are two events such that $P(A \cup B) = \frac{5}{6}$, $P(A \cap B) = \frac{1}{3}$, $P(\bar{A}) = \frac{1}{3}$,
the events are
- (i) dependent
 - (ii) independent
 - (iii) mutually exclusive
 - (iv) none of the above.
101. Which of the following statements are TRUE or FALSE :
- (i) The probability of an impossible event is always zero.
 - (ii) The number of permutations is always greater than the number of combinations.
 - (iii) If two events are independent, then they will also be mutually exclusive.
 - (iv) If $P(A)$ and $P(B)$ are non-zero and A and B are independent, then they cannot be mutually exclusive.
 - (v) If $P(A)$ and $P(B)$ are non-zero and A and B are mutually exclusive, then they may be independent.
 - (vi) The probability that the roof of a room will fall on the floor can be determined with the help of Classical definition.
 - (vii) Personal judgement or experience cannot be used in the assignment of probabilities.
 - (viii) Revision of the past probabilities of various events is possible on the basis of the outcome of the experiment.
 - (ix) The probability of occurrence of an event cannot be a negative number.
 - (x) The probability of occurrence of an event that is sure to occur can be greater than unity.
102. Objective Type Questions :
- (a) The probability of getting a number greater than 4 from the throw of an unbiased dice is

- (i) $\frac{1}{2}$ (ii) $\frac{1}{3}$ (iii) $\frac{1}{4}$ (iv) none of these.
- (b) The probability of getting exactly one tail in the toss of two unbiased coins is
 (i) $\frac{1}{2}$ (ii) $\frac{2}{3}$ (iii) $\frac{1}{4}$ (iv) none of these.
- (c) If odds in favour of an event A are $3 : 5$, then the probability of non-occurrence of A is
 (i) $\frac{3}{5}$ (ii) $\frac{3}{8}$ (iii) $\frac{5}{8}$ (iv) none of these.
- (d) Four dice and six coins are tossed simultaneously. The number of elements in the sample space are
 (i) $4^6 \times 6^2$ (ii) $2^6 \times 6^2$ (iii) $6^4 \times 2^6$ (iv) none of these.
- (e) Two cards are drawn successively without replacement from a well-shuffled pack of 52 cards. The probability that one of them is king and the other is queen is
 (i) $\frac{8}{13 \times 51}$ (ii) $\frac{4}{13 \times 51}$ (iii) $\frac{1}{13 \times 17}$ (iv) none of these.
- (f) Two unbiased dice are rolled. The chance of obtaining an even sum is
 (i) $\frac{1}{4}$ (ii) $\frac{1}{2}$ (iii) $\frac{1}{3}$ (iv) none of these.
- (g) Two unbiased dice are rolled. The chance of obtaining a six only on the second die is
 (i) $\frac{5}{6}$ (ii) $\frac{1}{6}$ (iii) $\frac{1}{4}$ (iv) none of these.
- (h) If $P(A) = \frac{4}{5}$, then odds against \bar{A} are
 (i) $1 : 4$ (ii) $5 : 4$ (iii) $4 : 5$ (iv) none of these.
- (i) The probability of occurrence of an event A is 0.60 and that of B is 0.25. If A and B are mutually exclusive events, then the probability of occurrence of neither of them is
 (i) 0.35 (ii) 0.75 (iii) 0.15 (iv) none of these.
- (j) The probability of getting at least one head in 3 throws of an unbiased coin is
 (i) $\frac{1}{8}$ (ii) $\frac{7}{8}$ (iii) $\frac{3}{8}$ (iv) none of these.

10.13 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

1. (a) Statistical, Random (b) Permutation (c) enumeration
 (d) Russian (e) Inverse Probability
2. (a) True (b) True (c) True (d) False (e) True

10.14 SUGGESTED READINGS

W. Feller, *An introduction to probability theory and its applications*, Volume 1, John Willy & Sons.

I Csiszar, *I divergence geometry of probability distributions and minimization problems*, The Annals of probability, Vol 3, No 1, pp. 146-158.

Sheldon M. Ross, *A first course in probability*, Prentice Hall.

Morris H. Degroot, Mark J, Schervish, "Probability and Statistics," Addison-Wesley, 2001

William Mendenhall, Robert, J. Beaver, Barbara, M. Beaver, *Introduction to Probability and Statistics*.

LESSON

11

THEORETICAL PROBABILITY DISTRIBUTIONS

CONTENTS

- 11.0 Aims and Objectives
- 11.1 Introduction
- 11.2 Probability Distribution
- 11.3 Binomial Distribution
- 11.4 Hypergeometric Distribution
- 11.5 Pascal Distribution
- 11.6 Geometrical Distribution
- 11.7 Uniform Distribution (Discrete Random Variable)
- 11.8 Poisson Distribution
- 11.9 Exponential Distribution
- 11.10 Uniform Distribution (Continuous Variable)
- 11.11 Normal Distribution
- 11.12 Let us Sum Up
- 11.13 Lesson-end Activity
- 11.14 Keywords
- 11.15 Questions for Discussion
- 11.16 Terminal Questions
- 11.17 Model Answers to Questions for Discussion
- 11.18 Suggested Readings

11.0 AIMS AND OBJECTIVES

After studying the probability in the previous lesson the next step is the probability distributions. Probability distributions are fundamental concepts of statistics. It is used both in theoretical aspects as well as practical aspects.

11.1 INTRODUCTION

Usual manager is forced to make decisions when there is uncertainty as to what will happen after the decisions are made. In this situation the mathematical theory of probability furnishes a tool that can be of great help to the decision maker. A probability function is a rule that assigns probabilities to each element of a set of events that may occur. Probability distribution can either discrete or continuous. A discrete probability distribution is sometimes called a probability mass function and a continuous one is called a probability density function.

11.2 PROBABILITY DISTRIBUTION

The study of a population can be done either by constructing an observed (or empirical) frequency distribution, often based on a sample from it, or by using a theoretical distribution. We have already studied the construction of an observed frequency distribution and its various summary measures. Now we shall learn a more scientific way to study a population through the use of theoretical probability distribution of a random variable. It may be mentioned that a theoretical probability distribution gives us a law according to which different values of the random variable are distributed with specified probabilities. It is possible to formulate such laws either on the basis of given conditions (a prior considerations) or on the basis of the results (a posteriori inferences) of an experiment.

If a random variable satisfies the conditions of a theoretical probability distribution, then this distribution can be fitted to the observed data.

The knowledge of the theoretical probability distribution is of great use in the understanding and analysis of a large number of business and economic situations. For example, with the use of probability distribution, it is possible to test a hypothesis about a population, to take decision in the face of uncertainty, to make forecast, etc.

Theoretical probability distributions can be divided into two broad categories, viz. discrete and continuous probability distributions, depending upon whether the random variable is discrete or continuous. Although, there are a large number of distributions in each category, we shall discuss only some of them having important business and economic applications.

11.3 BINOMIAL DISTRIBUTION

Binomial distribution is a theoretical probability distribution which was given by James Bernoulli. This distribution is applicable to situations with the following characteristics :

1. An experiment consists of a finite number of repeated trials.
2. Each trial has only two possible, mutually exclusive, outcomes which are termed as a 'success' or a 'failure'.
3. The probability of a success, denoted by p , is known and remains constant from trial to trial. The probability of a failure, denoted by q , is equal to $1 - p$.
4. Different trials are independent, i.e., outcome of any trial or sequence of trials has no effect on the outcome of the subsequent trials.

The sequence of trials under the above assumptions is also termed as *Bernoulli Trials*.

Probability Function or Probability Mass Function

Let n be the total number of repeated trials, p be the probability of a success in a trial and q be the probability of its failure so that $q = 1 - p$.

Let r be a random variable which denotes the number of successes in n trials. The possible values of r are $0, 1, 2, \dots, n$. We are interested in finding the probability of r successes out of n trials, i.e., $P(r)$.

To find this probability, we assume that the first r trials are successes and remaining $n - r$ trials are failures. Since different trials are assumed to be independent, the probability of this sequence is

$$\underbrace{p \cdot p \cdot \dots \cdot p}_{r \text{ times}} \underbrace{q \cdot q \cdot \dots \cdot q}_{(n-r) \text{ times}} \text{ i.e. } p^r q^{n-r}.$$

Since out of n trials any r trials can be success, the number of sequences showing any r trials as success and remaining $(n - r)$ trials as failure is ${}^n C_r$, where the probability of r successes in each trial is $p^r q^{n-r}$. Hence, the required probability is $P(r) = {}^n C_r p^r q^{n-r}$, where $r = 0, 1, 2, \dots, n$.

Writing this distribution in a tabular form, we have

r	0	1	2	n	Total
$P(r)$	${}^n C_0 p^0 q^n$	${}^n C_1 p^1 q^{n-1}$	${}^n C_2 p^2 q^{n-2}$	${}^n C_n p^n q^0$	1

It should be noted here that the probabilities obtained for various values of r are the terms in the binomial expansion of $(q + p)^n$ and thus, the distribution is termed as Binomial Distribution. $P(r) = {}^n C_r p^r q^{n-r}$ is termed as the probability function or probability mass function (p.m.f.) of the distribution.

Summary Measures of Binomial Distribution

(a) **Mean:** The mean of a binomial variate r , denoted by μ , is equal to $E(r)$, i.e.,

$$\begin{aligned}\mu = E(r) &= \sum_{r=0}^n r P(r) = \sum_{r=1}^n r \cdot {}^n C_r p^r q^{n-r} \quad (\text{note that the term for } r=0 \text{ is 0}) \\ &= \sum_{r=1}^n \frac{r \cdot n!}{r!(n-r)!} \cdot p^r q^{n-r} = \sum_{r=1}^n \frac{n(n-1)!}{(r-1)!(n-r)!} \cdot p^r q^{n-r} \\ &= np \sum_{r=1}^n \frac{(n-1)!}{(r-1)!(n-r)!} \cdot p^{r-1} q^{n-r} = np(q+p)^{n-1} = np \quad [\because q+p=1]\end{aligned}$$

(b) **Variance:** The variance of r , denoted by σ^2 , is given by

$$\begin{aligned}\sigma^2 &= E[r - E(r)]^2 = E[r - np]^2 = E[r^2 - 2np(r) + np^2] \\ &= E(r^2) - 2npE(r) + np^2 = E(r^2) - 2n^2p^2 + np^2 \\ &= E(r^2) - n^2p^2 \quad \dots (1)\end{aligned}$$

Thus, to find σ^2 , we first determine $E(r^2)$.

$$\begin{aligned}\text{Now, } E(r^2) &= \sum_{r=1}^n r^2 \cdot {}^n C_r p^r q^{n-r} = [r(r-1)+r] {}^n C_r p^r q^{n-r} \\ &= \sum_{r=2}^n r(r-1) {}^n C_r p^r q^{n-r} + \sum_{r=1}^n r \cdot {}^n C_r p^r q^{n-r} = \sum_{r=2}^n \frac{r(r-1)n!}{r!(n-r)!} \cdot p^r q^{n-r} + np \\ &= \sum_{r=2}^n \frac{n!}{(r-2)!(n-r)!} \cdot p^r q^{n-r} + np = \sum_{r=2}^n \frac{n(n-1)(n-2)!}{(r-2)!(n-r)!} \cdot p^r q^{n-r} + np \\ &= n(n-1)p^2 \sum_{r=2}^n \frac{(n-2)!}{(r-2)!(n-r)!} \cdot p^{r-2} q^{n-r} + np \\ &= n(n-1)p^2 (q+p)^{n-2} + np = n(n-1)p^2 + np\end{aligned}$$

Substituting this value in equation (1), we get

$$\sigma^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq$$

Or the standard deviation = \sqrt{npq}

Remarks: $\sigma^2 = npq = \text{mean} \times q$, which shows that $\sigma^2 < \text{mean}$, since $0 < q < 1$.

- (c) The values of μ_3 , μ_4 , β_1 and β_2

Proceeding as above, we can obtain

$$\mu_3 = E(r - np)^3 = npq(q - p)$$

$$\mu_4 = E(r - np)^4 = 3n^2 p^2 q^2 + npq(1 - 6pq)$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^2} = \frac{n^2 p^2 q^2 (q - p)^2}{n^3 p^3 q^3} = \frac{(q - p)^2}{npq}$$

Also

The above result shows that the distribution is symmetrical when

$p = q = \frac{1}{2}$, negatively skewed if $q < p$, and positively skewed if $q > p$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3n^2 p^2 q^2 + npq(1 - 6pq)}{n^2 p^2 q^2} = 3 + \frac{(1 - 6pq)}{npq}$$

The above result shows that the distribution is leptokurtic if $6pq < 1$, platykurtic if $6pq > 1$ and mesokurtic if $6pq = 1$.

- (d) **Mode:** Mode is that value of the random variable for which probability is maximum.

If r is mode of a binomial distribution, then we have

$$P(r - 1) \leq P(r) \geq P(r + 1)$$

Consider the inequality $P(r) \geq P(r + 1)$

$$\text{or } {}^n C_r p^r q^{n-r} \geq {}^n C_{r+1} p^{r+1} q^{n-r-1}$$

$$\text{or } \frac{n!}{r!(n-r)!} p^r q^{n-r} \geq \frac{n!}{(r+1)!(n-r-1)!} p^{r+1} q^{n-r-1}$$

$$\text{or } \frac{1}{(n-r)} \cdot q \geq \frac{1}{(r+1)} \cdot p \text{ or } qr + q \geq np - pr$$

Solving the above inequality for r , we get

$$r \geq (n+1)p - 1 \quad \dots (1)$$

Similarly, on solving the inequality $P(r - 1) \leq P(r)$ for r , we can get

$$r \leq (n+1)p \quad \dots (2)$$

Combining inequalities (1) and (2), we get

$$(n+1)p - 1 \leq r \leq (n+1)p$$

Case I: When $(n+1)p$ is not an integer

When $(n+1)p$ is not an integer, then $(n+1)p - 1$ is also not an integer. Therefore, mode will be an integer between $(n+1)p - 1$ and $(n+1)p$ or mode will be an integral part of $(n+1)p$.

Case II: When $(n+1)p$ is an integer

When $(n+1)p$ is an integer, the distribution will be bimodal and the two modal values would be $(n+1)p - 1$ and $(n+1)p$.

Example 1: An unbiased die is tossed three times. Find the probability of obtaining (a) no six, (b) one six, (c) at least one six, (d) two sixes and (e) three sixes.

Solution: The three tosses of a die can be taken as three repeated trials which are independent. Let the occurrence of six be termed as a success. Therefore, r will denote

the number of six obtained. Further, $n = 3$ and $p = \frac{1}{6}$.

(a) Probability of obtaining no six, i.e.,

$$P(r=0) = {}^3C_0 p^0 q^3 = 1 \cdot \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^3 = \frac{125}{216}$$

$$(b) P(r=1) = {}^3C_1 p^1 q^2 = 3 \cdot \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2 = \frac{25}{72}$$

$$(c) \text{Probability of getting at least one six} = 1 - P(r=0) = 1 - \frac{125}{216} = \frac{91}{216}$$

$$(d) P(r=2) = {}^3C_2 p^2 q^1 = 3 \cdot \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right) = \frac{5}{72}$$

$$(e) P(r=3) = {}^3C_3 p^3 q^0 = 3 \cdot \left(\frac{1}{6}\right)^3 = \frac{1}{216}$$

Example 2: Assuming that it is true that 2 in 10 industrial accidents are due to fatigue, find the probability that:

(a) Exactly 2 of 8 industrial accidents will be due to fatigue.

(b) At least 2 of the 8 industrial accidents will be due to fatigue.

Solution: Eight industrial accidents can be regarded as Bernoulli trials each with probability of success $p = \frac{2}{10} = \frac{1}{5}$. The random variable r denotes the number of accidents due to fatigue.

$$(a) P(r=2) = {}^8C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^6 = 0.294$$

(b) We have to find $P(r \geq 2)$. We can write

$$P(r \geq 2) = 1 - P(0) - P(1), \text{ thus, we first find } P(0) \text{ and } P(1).$$

$$\text{We have } P(0) = {}^8C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^8 = 0.168$$

$$\text{and } P(1) = {}^8C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^7 = 0.336$$

$$\therefore P(r \geq 2) = 1 - 0.168 - 0.336 = 0.496$$

Example 3: The proportion of male and female students in a class is found to be 1 : 2. What is the probability that out of 4 students selected at random with replacement, 2 or more will be females?

Solution: Let the selection of a female student be termed as a success. Since the selection of a student is made with replacement, the selection of 4 students can be taken as 4

repeated trials each with probability of success $p = \frac{2}{3}$.

$$\text{Thus, } P(r \geq 2) = P(r=2) + P(r=3) + P(r=4)$$

$$= {}^4C_2 \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^2 + {}^4C_3 \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right) + {}^4C_4 \left(\frac{2}{3}\right)^4 = \frac{8}{9}$$

Note that $P(r \geq 2)$ can alternatively be found as $1 - P(0) - P(1)$

Example 4: The probability of a bomb hitting a target is $1/5$. Two bombs are enough to destroy a bridge. If six bombs are aimed at the bridge, find the probability that the bridge is destroyed.

Solution: Here $n = 6$ and $p = \frac{1}{5}$

The bridge will be destroyed if at least two bombs hit it. Thus, we have to find $P(r \geq 2)$. This is given by

$$P(r \geq 2) = 1 - P(0) - P(1) = 1 - {}^6C_0 \left(\frac{4}{5}\right)^6 - {}^6C_1 \left(\frac{1}{5}\right) \left(\frac{4}{5}\right)^5 = \frac{1077}{3125}$$

Example 5: An insurance salesman sells policies to 5 men all of identical age and good health. According to the actuarial tables, the probability that a man of this particular age will be alive 30 years hence is $2/3$. Find the probability that 30 years hence (i) at least 1 man will be alive, (ii) at least 3 men will be alive.

Solution: Let the event that a man will be alive 30 years hence be termed as a success.

Therefore, $n = 5$ and $p = \frac{2}{3}$.

$$(i) \quad P(r \geq 1) = 1 - P(r = 0) = 1 - {}^5C_0 \left(\frac{2}{3}\right)^0 \left(\frac{1}{3}\right)^5 = \frac{242}{243}$$

$$(ii) \quad P(r \geq 3) = P(r = 3) + P(r = 4) + P(r = 5)$$

$$= {}^5C_3 \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^2 + {}^5C_4 \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right) + {}^5C_5 \left(\frac{2}{3}\right)^5 = \frac{64}{81}$$

Example 6: Ten percent of items produced on a machine are usually found to be defective. What is the probability that in a random sample of 12 items (i) none, (ii) one, (iii) two, (iv) at the most two, (v) at least two items are found to be defective?

Solution: Let the event that an item is found to be defective be termed as a success. Thus, we are given $n = 12$ and $p = 0.1$.

$$(i) \quad P(r = 0) = {}^{12}C_0 (0.1)^0 (0.9)^{12} = 0.2824$$

$$(ii) \quad P(r = 1) = {}^{12}C_1 (0.1)^1 (0.9)^{11} = 0.3766$$

$$(iii) \quad P(r = 2) = {}^{12}C_2 (0.1)^2 (0.9)^{10} = 0.2301$$

$$(iv) \quad P(r \leq 2) = P(r = 0) + P(r = 1) + P(r = 2) \\ = 0.2824 + 0.3766 + 0.2301 = 0.8891$$

$$(v) \quad P(r \geq 2) = 1 - P(0) - P(1) = 1 - 0.2824 - 0.3766 = 0.3410$$

Example 7: In a large group of students 80% have a recommended statistics book. Three students are selected at random. Find the probability distribution of the number of students having the book. Also compute the mean and variance of the distribution.

Solution: Let the event that 'a student selected at random has the book' be termed as a success. Since the group of students is large, 3 trials, i.e., the selection of 3 students, can be regarded as independent with probability of a success $p = 0.8$. Thus, the conditions of the given experiment satisfies the conditions of binomial distribution.

The probability mass function $P(r) = {}^3C_r (0.8)^r (0.2)^{3-r}$,

where $r = 0, 1, 2$ and 3

The mean is $np = 3 \times 0.8 = 2.4$ and Variance is $npq = 2.4 \times 0.2 = 0.48$

Example 8:

- (a) The mean and variance of a discrete random variable X are 6 and 2 respectively. Assuming X to be a binomial variate, find $P(5 \leq X \leq 7)$.
- (b) In a binomial distribution consisting of 5 independent trials, the probability of 1 and 2 successes are 0.4096 and 0.2048 respectively. Calculate the mean, variance and mode of the distribution.

Solution:

- (a) It is given that $np = 6$ and $npq = 2$

$$\therefore q = \frac{npq}{np} = \frac{2}{6} = \frac{1}{3} \text{ so that } p = 1 - \frac{1}{3} = \frac{2}{3} \text{ and } n = 6 \times \frac{3}{2} = 9$$

Now $P(5 \leq X \leq 7) = P(X = 5) + P(X = 6) + P(X = 7)$

$$\begin{aligned} &= {}^9C_5 \left(\frac{2}{3}\right)^5 \left(\frac{1}{3}\right)^4 + {}^9C_6 \left(\frac{2}{3}\right)^6 \left(\frac{1}{3}\right)^3 + {}^9C_7 \left(\frac{2}{3}\right)^7 \left(\frac{1}{3}\right)^2 \\ &= \frac{2^5}{3^9} \left[{}^9C_5 + {}^9C_6 \times 2 + {}^9C_7 \times 4 \right] = \frac{2^5}{3^9} \times 438 \end{aligned}$$

- (b) Let p be the probability of a success. It is given that

$${}^5C_1 p (1-p)^4 = 0.4096 \text{ and } {}^5C_2 p^2 (1-p)^3 = 0.2048$$

Using these conditions, we can write

$$\frac{5p(1-p)^4}{10p^2(1-p)^3} = \frac{0.4096}{0.2048} = 2 \text{ or } \frac{(1-p)}{p} = 4. \text{ This gives } p = \frac{1}{5}$$

Thus, mean is $np = 5 \times \frac{1}{5} = 1$ and $npq = 1 \times \frac{4}{5} = 0.8$

Since $(n+1)p$, i.e., $6 \times \frac{1}{5}$ is not an integer, mode is its integral part, i.e., = 1.

Example 9: 5 unbiased coins are tossed simultaneously and the occurrence of a head is termed as a success. Write down various probabilities for the occurrence of 0, 1, 2, 3, 4, 5 successes. Find mean, variance and mode of the distribution.

Solution: Here $n = 5$ and $p = q = \frac{1}{2}$.

The probability mass function is $P(r) = {}^5C_r \left(\frac{1}{2}\right)^5$, $r = 0, 1, 2, 3, 4, 5$.

The probabilities of various values of r are tabulated below :

r	0	1	2	3	4	5	Total
$P(r)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$	1

$$\text{Mean} = np = 5 \times \frac{1}{2} = 2.5 \text{ and variance} = 2.5 \times \frac{1}{2} = 1.25$$

Since $(n+1)p = 6 \times \frac{1}{2} = 3$ is an integer, the distribution is bimodal and the two modes are 2 and 3.

Fitting of Binomial Distribution

The fitting of a distribution to given data implies the determination of expected (or theoretical) frequencies for different values of the random variable on the basis of this data.

The purpose of fitting a distribution is to examine whether the observed frequency distribution can be regarded as a sample from a population with a known probability distribution.

To fit a binomial distribution to the given data, we find its mean. Given the value of n, we can compute the value of p and, using n and p, the probabilities of various values of the random variable can be computed. These probabilities are multiplied by total frequency to give the required expected frequencies. In certain cases, the value of p may be determined by the given conditions of the experiment.

Example 10: The following data give the number of seeds germinating (X) out of 10 on damp filter for 80 sets of seed. Fit a binomial distribution to the data.

X :	0	1	2	3	4	5	6	7	8	9	10
f :	6	20	28	12	8	6	0	0	0	0	0

Solution: Here the random variable X denotes the number of seeds germinating out of a set of 10 seeds. The total number of trials n = 10.

The mean of the given data

$$\bar{X} = \frac{0 \times 6 + 1 \times 20 + 2 \times 28 + 3 \times 12 + 4 \times 8 + 5 \times 6}{80} = \frac{174}{80} = 2.175$$

Since mean of a binomial distribution is np, $\therefore np = 2.175$. Thus, we get .

$$p = \frac{2.175}{10} = 0.22 \text{ (approx.)}. \text{ Further, } q = 1 - 0.22 = 0.78.$$

Using these values, we can compute $P(X) = {}^{10}C_x (0.22)^x (0.78)^{10-x}$ and then expected frequency [= N × P(X)] for X = 0, 1, 2, 10. The calculated probabilities and the respective expected frequencies are shown in the following table:

X	P(X)	N×P(X)	Approximated Frequency	X	P(X)	N×P(X)	Approximated Frequency
0	0.0834	6.67	6	6	0.0088	0.71	1
1	0.2351	18.81	19	7	0.0014	0.11	0
2	0.2984	23.87	24	8	0.0001	0.01	0
3	0.2244	17.96	18	9	0.0000	0.00	0
4	0.1108	8.86	9	10	0.0000	0.00	0
5	0.0375	3.00	3	Total	1.0000		80

Features of Binomial Distribution

1. It is a discrete probability distribution.
2. It depends upon two parameters n and p. It may be pointed out that *a distribution is known if the values of its parameters are known.*
3. The total number of possible values of the random variable are n + 1. The successive binomial coefficients are ${}^nC_0, {}^nC_1, {}^nC_2, \dots, {}^nC_n$. Further, since ${}^nC_r = {}^nC_{n-r}$, these coefficients are symmetric.

The values of these coefficients, for various values of n, can be obtained directly by using Pascal's triangle.

PASCAL'S TRIANGLE

<i>n</i>	<i>Binomial Coefficients</i>	<i>Sum of Coefficients (2ⁿ)</i>
1	1 1	$2^1 = 2$
2	1 2 1	$2^2 = 4$
3	1 3 3 1	$2^3 = 8$
4	1 4 6 4 1	$2^4 = 16$
5	1 5 10 10 5 1	$2^5 = 32$

We can note that it is very easy to write this triangle. In the first row, both the coefficients will be unity because ${}^1C_0 = {}^1C_1$. To write the second row, we write 1 in the beginning and the end and the value of the middle coefficients is obtained by adding the coefficients of the first row. Other rows of the Pascal's triangle can be written in a similar way.

4. (a) The shape and location of binomial distribution changes as the value of p changes for a given value of n. It can be shown that for a given value of n, if p is increased gradually in the interval (0, 0.5), the distribution changes from a positively skewed to a symmetrical shape. When p = 0.5, the distribution is perfectly symmetrical. Further, for larger values of p the distribution tends to become more and more negatively skewed.
- (b) For a given value of p, which is neither too small nor too large, the distribution becomes more and more symmetrical as n becomes larger and larger.

Uses of Binomial Distribution

Binomial distribution is often used in various decision-making situations in business. Acceptance sampling plan, a technique of quality control, is based on this distribution. With the use of sampling plan, it is possible to accept or reject a lot of items either at the stage of its manufacture or at the stage of its purchase.

11.4 HYPERGEOMETRIC DISTRIBUTION

The binomial distribution is not applicable when the probability of a success p does not remain constant from trial to trial. In such a situation the probabilities of the various values of r are obtained by the use of Hypergeometric distribution.

Let there be a finite population of size N, where each item can be classified as either a success or a failure. Let there be k successes in the population. If a random sample of size n is taken from this population, then the probability of r successes is given by

$$P(r) = \frac{{}^k C_r \left({}^{N-k} C_{n-r} \right)}{{}^N C_n}. \text{ Here } r \text{ is a discrete random variable which can take values}$$

0, 1, 2, ..., n. Also $n \leq k$.

It can be shown that the mean of r is np and its variance is

$$\left(\frac{N-n}{N-1} \right) npq, \text{ where } p = \frac{k}{N} \text{ and } q = 1 - p.$$

Example 11: A retailer has 10 identical television sets of a company out of which 4 are defective. If 3 televisions are selected at random, construct the probability distribution of the number of defective television sets.

Solution: Let the random variable r denote the number of defective televisions. In terms of notations, we can write N = 10, k = 4 and n = 3.

Thus, we can write $P(r) = \frac{^4C_r \times ^6C_{3-r}}{^{10}C_3}$, $r = 0, 1, 2, 3$

The distribution of r is hypergeometric. This distribution can also be written in a tabular form as given below :

r	0	1	2	3	Total
$P(r)$	5	15	9	1	1
	30	30	30	30	

Binomial Approximation to Hypergeometric Distribution

In sampling problems, where sample size n (total number of trials) is less than 5% of population size N , i.e., $n < 0.05N$, the use of binomial distribution will also give satisfactory results. The reason for this is that the smaller the sample size relative to population size, the greater will be the validity of the requirements of independent trials and the constancy of p .

Example 12: There are 200 identical radios out of which 80 are defective. If 5 radios are selected at random, construct the probability distribution of the number of defective radios by using (i) hypergeometric distribution and (ii) binomial distribution.

Solution:

(i) It is given that $N = 200$, $k = 80$ and $n = 5$.

Let r be a hypergeometric random variable which denotes the number of defective radios, then

$$P(r) = \frac{^{80}C_r \times ^{120}C_{5-r}}{^{200}C_5}, \quad r = 0, 1, 2, 3, 4, 5$$

The probabilities for various values of r are given in the following table :

r	0	1	2	3	4	5	Total
$P(r)$	0.0752	0.2592	0.3500	0.2313	0.0748	0.0095	1

(ii) To use binomial distribution, we find $p = 80/200 = 0.4$.

$$P(r) = {}^5C_r (0.4)^r (0.6)^{5-r}, \quad r = 0, 1, 2, 3, 4, 5$$

The probabilities for various values of r are given in the following table :

r	0	1	2	3	4	5	Total
$P(r)$	0.0778	0.2592	0.3456	0.2304	0.0768	0.0102	1

We note that these probabilities are in close conformity with the hypergeometric probabilities.

Check Your Progress 11.1

1. Write the characteristics of Binomial Distribution.
2. What is the use of Hypergeometric Distribution?

Notes: (a) Write your answer in the space given below.

(b) Please go through the lesson sub-head thoroughly you will get your answers in it.

- (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-
-

11.5 PASCAL DISTRIBUTION

In binomial distribution, we derived the probability mass function of the number of successes in n (fixed) Bernoulli trials. We can also derive the probability mass function of the number of Bernoulli trials needed to get r (fixed) successes. This distribution is known as Pascal distribution. Here r and p become parameters while n becomes a random variable.

We may note that r successes can be obtained in r or more trials i.e. possible values of the random variable are r, (r + 1), (r + 2), etc. Further, if n trials are required to get r successes, the nth trial must be a success. Thus, we can write the probability mass function of Pascal distribution as follows :

$$P(n) = \binom{\text{Probability of } (r-1) \text{ successes}}{\text{out of } (n-1) \text{ trials}} \times \binom{\text{Probability of a success}}{\text{in } n\text{th trial}}$$

$$= {}^{n-1}C_{r-1} p^{r-1} q^{n-r} \times p = {}^{n-1}C_{r-1} p^r q^{n-r}, \text{ where } n = r, (r+1), (r+2), \dots \text{ etc.}$$

It can be shown that the mean and variance of Pascal distribution are $\frac{r}{p}$ and $\frac{rq}{p^2}$ respectively.

This distribution is also known as *Negative Binomial Distribution* because various values of P(n) are given by the terms of the binomial expansion of $p^r(1 - q)^{-r}$.

11.6 GEOMETRICAL DISTRIBUTION

When r = 1, the Pascal distribution can be written as

$$P(n) = {}^{n-1}C_0 p q^{n-1} = p q^{n-1}, \text{ where } n = 1, 2, 3, \dots$$

Here n is a random variable which denotes the number of trials required to get a success. This distribution is known as geometrical distribution. The mean and variance of the

distribution are $\frac{1}{p}$ and $\frac{q}{p^2}$ respectively.

11.7 UNIFORM DISTRIBUTION (DISCRETE RANDOM VARIABLE)

A discrete random variable is said to follow a uniform distribution if it takes various discrete values with equal probabilities.

If a random variable X takes values X_1, X_2, \dots, X_n each with probability $\frac{1}{n}$, the distribution of X is said to be uniform.

Exercise with Hints

1. The probability that a secretary will not put the correct postage on a letter is 0.20.
What is the probability that this secretary will not put the correct postage:
(i) On 3 of 9 letters? (ii) On at least 3 of 9 letters? (iii) On at the most 3 of 9 letters?

Hint: Use binomial distribution.

2. (a) The mean of a binomial distribution is 4 and its standard deviation is $\sqrt{3}$. What are the values of n, p and q?
(b) The mean and variance of a binomial distribution are 3 and 2 respectively. Find the probability that the variate takes values (i) less than or equal to 2 (ii) greater than or equal to 7.

Hint: Mean = np and Variance = npq.

3. (a) The probability of a man hitting a target is $\frac{1}{4}$. (i) If he fires 7 times, what is the probability of his hitting the target at least twice? (ii) How many times must he fire so that the probability of his hitting the target at least once is greater than $\frac{2}{3}$?
(b) How many dice must be thrown so that there is better than even chance of obtaining at least one six?

Hint: (a) (ii) Probability of hitting the target at least once in n trials is $1 - \left(\frac{3}{4}\right)^n$.

Find n such that this value is greater than $\frac{2}{3}$. (b) Find n so that

$$1 - \left(\frac{5}{6}\right)^n > \frac{1}{2}.$$

4. A machine produces an average of 20% defective bolts. A batch is accepted if a sample of 5 bolts taken from the batch contains no defective and rejected if the sample contains 3 or more defectives. In other cases, a second sample is taken. What is the probability that the second sample is required?

Hint: A second sample is required if the first sample is neither rejected nor accepted.

5. A multiple choice test consists of 8 questions with 3 answers to each question (of which only one is correct). A student answers each question by throwing a balanced die and checking the first answer if he gets 1 or 2, the second answer if he gets 3 or 4 and the third answer if he gets 5 or 6. To get a distinction, the student must secure at least 75% correct answers. If there is no negative marking, what is the probability that the student secures a distinction?

Hint: He should attempt at least 6 questions.

6. What is the most probable number of times an ace will appear if a die is tossed (i) 50 times, (ii) 53 times?

Hint: Find mode.

7. Out of 320 families with 5 children each, what percentage would be expected to have (i) 2 boys and 3 girls, (ii) at least one boy, (iii) at the most one girl? Assume equal probability for boys and girls.

Hint: Multiply probability by 100 to obtain percentage.

8. Fit a binomial distribution to the following data :

X :	0	1	2	3	4
f :	28	62	46	10	4

Hint: See example 10.

9. A question paper contains 6 questions of equal value divided into two sections of three questions each. If each question poses the same amount of difficulty to Mr. X, an examinee, and he has only 50% chance of solving it correctly, find the answer to the following :

- (i) If Mr. X is required to answer only three questions from any one of the two sections, find the probability that he will solve all the three questions correctly.
- (ii) If Mr. X is given the option to answer the three questions by selecting one question out of the two standing at serial number one in the two sections, one question out of the two standing at serial number two in the two sections and one question out of the two standing at serial number three in the two sections, find the probability that he will solve all three questions correctly.

Hint: (i) A section can be selected in 2C_1 ways and the probability of attempting all the three questions correctly is ${}^3C_3 \left(\frac{1}{2}\right)^3$. (ii) The probability of attempting correctly, one question out of two is ${}^2C_1 \left(\frac{1}{2}\right)^2$

10. A binomial random variable satisfies the relation $9P(X = 4) = P(X = 2)$ for $n = 6$. Find the value of the parameter p.

Hint: $P(X = 2) = {}^6C_2 p^2 q^4$ etc.

11. Three fair coins are tossed 3,000 times. Find the frequency distributions of the number of heads and tails and tabulate the results. Also calculate mean and standard deviation of each distribution.

Hint: See example 9.

12. Take 100 sets of 10 tosses of an unbiased coin. In how many cases do you expect to get (i) 6 heads and 4 tails and (ii) at least 9 heads?

Hint: Use binomial distribution with $n = 10$ and $p = 0.5$.

13. In a binomial distribution consisting of 5 independent trials, the probabilities of 1 and 2 successes are 0.4096 and 0.2048 respectively. Find the probability of success.

Hint: Use the condition $P(1) = 2P(2)$.

14. For a binomial distribution, the mean and variance are respectively 4 and 3. Calculate the probability of getting a nonzero value of this distribution.

Hint: Find $P(r \neq 0)$.

15. (a) There are 300, seemingly identical, tyres with a dealer. The probability of a tyre being defective is 0.3. If 2 tyres are selected at random, find the probability that there is non defective tyre.
 (b) If instead of 300 tyres the dealer had only 10 tyres out of which 3 are defective, find the probability that no tyre is defective in a random sample of 2 tyres.
 (c) Write down the probability distribution of the number of defectives in each case.

Hint: Use (a) binomial (b) hypergeometric distributions.

16. Write down the mean and variance of a binomial distribution with parameters n and p. If the mean and variance are 4 and $8/3$ respectively, find the values of n and p. State whether it is symmetric for these values?

Hint: Binomial distribution is symmetric when $p = 0.5$.

17. Evaluate k if $f(x) = k$, when $x = 1, 2, 3, 4, 5, 6$ and $= 0$ elsewhere, is a probability mass function. Find its mean and standard deviation.

Hint: $\sum f(x) = 1$.

18. If a die is thrown 6 times, calculate the probability that :

- (i) a score of 3 or less occurs on exactly 2 throws;
- (ii) a score of more than 2 occurs on exactly 3 throws;
- (iii) a score of 5 or less occurs at least once;
- (iv) a score of 2 or less occurs on at least 5 throws.

Hint: Use binomial distribution with $n = 6$ and a different value of p in each case.

19. If we take 1,280 sets each of 10 tosses of a fair coins, in how many sets should we expect to get 7 heads and 3 tails?

Hint: See example 9.

20. If a production unit is made up from 20 identical components and each component has a probability of 0.25 of being defective, what is the average number of defective components in a unit? Further, what is the probability that in a unit (i) less than 3 components are defective? (ii) exactly 3 components are defective?

Hint: Take $n = 20$ and $p = 0.25$.

21. It is known from the past experience that 80% of the students in a school do their home work. Find the probability that during a random check of 10 students

- (i) all have done their home work,
- (ii) at the most 2 have not done their home work,
- (iii) at least one has not done the home work.

Hint: Take $n = 10$ and $p = 0.8$.

22. There are 24 battery cells in a box containing 6 defective cells that are randomly mixed. A customer buys 3 cells. What is the probability that he gets one defective cell?

Hint: Use hypergeometric distribution.

23. There are 400 tyres in the stock of a wholesaler among which 40 tyres, having slight defects, are randomly mixed. A retailer purchases 6 tyres from this stock. What is the probability that he gets at least 4 non defective tyres?

Hint: n is less than 5% of N .

11.8 POISSON DISTRIBUTION

This distribution was derived by a noted mathematician, Simon D. Poisson, in 1837. He derived this distribution as a limiting case of binomial distribution, when the number of trials n tends to become very large and the probability of success in a trial p tends to become very small such that their product np remains a constant. This distribution is used as a model to describe the probability distribution of a random variable defined over a unit of time, length or space. For example, the number of telephone calls received per hour at a telephone exchange, the number of accidents in a city per week, the number of defects per meter of cloth, the number of insurance claims per year, the number breakdowns of machines at a factory per day, the number of arrivals of customers at a shop per hour, the number of typing errors per page etc.

Poisson Process

Let us assume that on an average 3 telephone calls are received per 10 minutes at a telephone exchange desk and we want to find the probability of receiving a telephone call in the next 10 minutes. In an effort to apply binomial distribution, we can divide the interval of 10 minutes into 10 intervals of 1 minute each so that the probability of receiving a telephone call (i.e., a success) in each minute (i.e., trial) becomes 3/10 (note that $p = m/n$, where m denotes mean). Thus, there are 10 trials which are independent, each with probability of success = 3/10. However, the main difficulty with this formulation is that, strictly speaking, these trials are not Bernoulli trials. One essential requirement of such trials, that each trial must result into one of the two possible outcomes, is violated here. In the above example, a trial, i.e. an interval of one minute, may result into 0, 1, 2, successes depending upon whether the exchange desk receives none, one, two, telephone calls respectively.

One possible way out is to divide the time interval of 10 minutes into a large number of small intervals so that the probability of receiving two or more telephone calls in an interval becomes almost zero. This is illustrated by the following table which shows that the probabilities of receiving two calls decreases sharply as the number of intervals are increased, keeping the average number of calls, 3 calls in 10 minutes in our example, as constant.

n	$P(\text{one call is received})$	$P(\text{two calls are received})$
10	0.3	0.09
100	0.03	0.0009
1,000	0.003	0.000009
10,000	0.0003	0.00000009

Using symbols, we may note that as n increases then p automatically declines in such a way that the mean m ($= np$) is always equal to a constant. Such a process is termed as a *Poisson Process*. The chief characteristics of Poisson process can be summarised as given below :

1. The number of occurrences in an interval is independent of the number of occurrences in another interval.
2. The expected number of occurrences in an interval is constant.
3. It is possible to identify a small interval so that the occurrence of more than one event, in any interval of this size, becomes extremely unlikely.

Probability Mass Function

The probability mass function (p.m.f.) of Poisson distribution can be derived as a limit of p.m.f. of binomial distribution when $n \rightarrow \infty$ such that m ($= np$) remains constant. Thus, we can write

$$\begin{aligned}
 P(r) &= \lim_{n \rightarrow \infty} {}^n C_r \left(\frac{m}{n} \right)^r \left(1 - \frac{m}{n} \right)^{n-r} = \lim_{n \rightarrow \infty} \frac{n!}{r!(n-r)!} \left(\frac{m}{n} \right)^r \left(1 - \frac{m}{n} \right)^{n-r} \\
 &= \frac{m^r}{r!} \cdot \lim_{n \rightarrow \infty} \left[n(n-1)(n-2) \dots (n-r+1) \cdot \frac{1}{n^r} \cdot \left(1 - \frac{m}{n} \right)^{n-r} \right] \\
 &= \frac{m^r}{r!} \cdot \lim_{n \rightarrow \infty} \left[\frac{\frac{n}{n} \left(1 - \frac{1}{n} \right) \left(1 - \frac{2}{n} \right) \dots \left(1 - \frac{(r-1)}{n} \right) \left(1 - \frac{m}{n} \right)^n}{\left(1 - \frac{m}{n} \right)^r} \right]
 \end{aligned}$$

$= \frac{m^r}{r!} \lim_{n \rightarrow \infty} \left(1 - \frac{m}{n}\right)^n$, since each of the remaining terms will tend to unity as

$$n \rightarrow \infty$$

$$= \frac{m^r \cdot e^{-m}}{r!}, \text{ since } \lim_{n \rightarrow \infty} \left(1 - \frac{m}{n}\right)^n = \lim_{n \rightarrow \infty} \left\{ \left(1 - \frac{m}{n}\right)^{\frac{n}{m}} \right\}^m = e^{-m}.$$

Thus, the probability mass function of Poisson distribution is

$$P(r) = \frac{e^{-m} \cdot m^r}{r!}, \text{ where } r = 0, 1, 2, \dots, \infty.$$

Here e is a constant with value = 2.71828... . Note that Poisson distribution is a discrete probability distribution with single parameter m.

$$\begin{aligned} \text{Total probability} &= \sum_{r=0}^{\infty} \frac{e^{-m} \cdot m^r}{r!} = e^{-m} \left(1 + \frac{m}{1!} + \frac{m^2}{2!} + \frac{m^3}{3!} + \dots \right) \\ &= e^{-m} \cdot e^m = 1. \end{aligned}$$

Summary Measures of Poisson Distribution

(a) **Mean:** The mean of a Poisson variate r is defined as

$$\begin{aligned} E(r) &= \sum_{r=0}^{\infty} r \cdot \frac{e^{-m} \cdot m^r}{r!} = e^{-m} \sum_{r=1}^{\infty} \frac{m^r}{(r-1)!} = e^{-m} \left[m + m^2 + \frac{m^3}{2!} + \frac{m^4}{3!} + \dots \right] \\ &= m e^{-m} \left[1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \dots \right] = m e^{-m} e^m = m \end{aligned}$$

(b) **Variance:** The variance of a Poisson variate is defined as

$$\text{Var}(r) = E(r - m)^2 = E(r^2) - m^2$$

$$\begin{aligned} \text{Now } E(r^2) &= \sum_{r=0}^{\infty} r^2 P(r) = \sum_{r=0}^{\infty} [r(r-1) + r] P(r) = \sum_{r=0}^{\infty} [r(r-1)] P(r) + \sum_{r=0}^{\infty} r P(r) \\ &= \sum_{r=2}^{\infty} [r(r-1)] \frac{e^{-m} \cdot m^r}{r!} + m = e^{-m} \sum_{r=2}^{\infty} \frac{m^r}{(r-2)!} + m \\ &= m + e^{-m} \left(m^2 + m^3 + \frac{m^4}{2!} + \frac{m^5}{3!} + \dots \right) \\ &= m + m^2 e^{-m} \left(1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \dots \right) = m + m^2 \end{aligned}$$

Thus, $\text{Var}(r) = m + m^2 - m^2 = m$.

Also standard deviation $\sigma = \sqrt{m}$.

(c) The values of μ_3 , μ_4 , β_1 and β_2

It can be shown that $\mu_3 = m$ and $\mu_4 = m + 3m^2$.

$$\therefore \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{m^2}{m^3} = \frac{1}{m}$$

Since m is a positive quantity, therefore, β_1 is always positive and hence the Poisson distribution is always positively skewed. We note that $\beta_1 \rightarrow 0$ as $\mu \rightarrow \infty$, therefore the distribution tends to become more and more symmetrical for large values of m .

Further, $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{m + 3m^2}{m^2} = 3 + \frac{1}{m} \rightarrow 3$ as $m \rightarrow \infty$. This result shows that the distribution becomes normal for large values of m .

- (d) **Mode:** As in binomial distribution, a Poisson variate r will be mode if

$$P(r-1) \leq P(r) \geq P(r+1)$$

The inequality $P(r-1) \leq P(r)$ can be written as

$$\frac{e^{-m} \cdot m^{r-1}}{(r-1)!} \leq \frac{e^{-m} \cdot m^r}{r!} \Rightarrow 1 \leq \frac{m}{r} \Rightarrow r \leq m \quad \dots (1)$$

Similarly, the inequality $P(r) \geq P(r+1)$ can be shown to imply that

$$r \geq m - 1 \quad \dots (2)$$

Combining (1) and (2), we can write $m - 1 \leq r \leq m$.

Case I: When m is not an integer

The integral part of m will be mode.

Case II: When m is an integer

The distribution is bimodal with values m and $m - 1$.

Example 13: The average number of customer arrivals per minute at a super bazaar is 2. Find the probability that during one particular minute (i) exactly 3 customers will arrive, (ii) at the most two customers will arrive, (iii) at least one customer will arrive.

Solution: It is given that $m = 2$. Let the number of arrivals per minute be denoted by the random variable r . The required probability is given by

$$(i) \quad P(r=3) = \frac{e^{-2} \cdot 2^3}{3!} = \frac{0.13534 \times 8}{6} = 0.18045$$

$$(ii) \quad P(r \leq 2) = \sum_{r=0}^2 \frac{e^{-2} \cdot 2^r}{r!} = e^{-2} \left[1 + 2 + \frac{4}{2} \right] = 0.13534 \times 5 = 0.6767.$$

$$(iii) \quad P(r \geq 1) = 1 - P(r=0) = 1 - \frac{e^{-2} \cdot 2^0}{0!} = 1 - 0.13534 = 0.86464.$$

Example 14: An executive makes, on an average, 5 telephone calls per hour at a cost which may be taken as Rs 2 per call. Determine the probability that in any hour the telephone calls' cost (i) exceeds Rs 6, (ii) remains less than Rs 10.

Solution: The number of telephone calls per hour is a random variable with mean = 5. The required probability is given by

$$\begin{aligned}
 \text{(i)} \quad P(r > 3) &= 1 - P(r \leq 3) = 1 - \sum_{r=0}^3 \frac{e^{-5} \cdot 5^r}{r!} \\
 &= 1 - e^{-5} \left[1 + 5 + \frac{25}{2} + \frac{125}{6} \right] = 1 - 0.00678 \times \frac{236}{6} = 0.7349. \\
 \text{(ii)} \quad P(r \leq 4) &= \sum_{r=0}^4 \frac{e^{-5} \cdot 5^r}{r!} = e^{-5} \left[1 + 5 + \frac{25}{2} + \frac{125}{6} + \frac{625}{24} \right] = 0.00678 \times \frac{1569}{24} = 0.44324.
 \end{aligned}$$

Example 15: A company makes electric toys. The probability that an electric toy is defective is 0.01. What is the probability that a shipment of 300 toys will contain exactly 5 defectives?

Solution: Since n is large and p is small, Poisson distribution is applicable. The random variable is the number of defective toys with mean $m = np = 300 \times 0.01 = 3$. The required probability is given by

$$P(r = 5) = \frac{e^{-3} \cdot 3^5}{5!} = \frac{0.04979 \times 243}{120} = 0.10082.$$

Example 16: In a town, on an average 10 accidents occur in a span of 50 days. Assuming that the number of accidents per day follow Poisson distribution, find the probability that there will be three or more accidents in a day.

Solution: The random variable denotes the number accidents per day. Thus, we have .

$$m = \frac{10}{50} = 0.2. \text{ The required probability is given by}$$

$$P(r \geq 3) = 1 - P(r \leq 2) = 1 - e^{-0.2} \left[1 + 0.2 + \frac{(0.2)^2}{2!} \right] = 1 - 0.8187 \times 1.22 = 0.00119.$$

Example 17: A car hire firm has two cars which it hire out every day. The number of demands for a car on each day is distributed as a Poisson variate with mean 1.5. Calculate the proportion of days on which neither car is used and the proportion of days on which some demand is refused. [$e^{-1.5} = 0.2231$]

Solution: When both car are not used, $r = 0$

$\therefore P(r = 0) = e^{-1.5} = 0.2231$. Hence the proportion of days on which neither car is used is 22.31%.

Further, some demand is refused when more than 2 cars are demanded, i.e., $r > 2$

$$\therefore P(r > 2) = 1 - P(r \leq 2) = 1 - \sum_{r=0}^2 \frac{e^{-1.5} (1.5)^r}{r!} = 1 - 0.2231 \left[1 + 1.5 + \frac{(1.5)^2}{2!} \right] = 0.1913.$$

Hence the proportion of days is 19.13%.

Example 18: A firm produces articles of which 0.1 percent are usually defective. It packs them in cases each containing 500 articles. If a wholesaler purchases 100 such cases, how many cases are expected to be free of defective items and how many are expected to contain one defective item?

Solution: The Poisson variate is number of defective items with mean

$$m = \frac{1}{1000} \times 500 = 0.5.$$

$P(r=0) = e^{-0.5} = 0.6065$. Hence the number of cases having no defective items $= 0.6065 \times 100 = 60.65$

Similarly, $P(r=1) = e^{-0.5} \times 0.5 = 0.6065 \times 0.5 = 0.3033$. Hence the number of cases having one defective item are 30.33.

Example 19: A manager accepts the work submitted by his typist only when there is no mistake in the work. The typist has to type on an average 20 letters per day of about 200 words each. Find the chance of her making a mistake (i) if less than 1% of the letters submitted by her are rejected; (ii) if on 90% of days all the work submitted by her is accepted. [As the probability of making a mistake is small, you may use Poisson distribution. Take $e = 2.72$].

Solution: Let p be the probability of making a mistake in typing a word.

- (i) Let the random variable r denote the number of mistakes per letter. Since 20 letters are typed, r will follow Poisson distribution with mean $= 20p$.

Since less than 1% of the letters are rejected, it implies that the probability of making at least one mistake is less than 0.01, i.e.,

$$P(r \geq 1) \leq 0.01 \text{ or } 1 - P(r = 0) \leq 0.01$$

$$\Rightarrow 1 - e^{-20p} \leq 0.01 \text{ or } e^{-20p} \leq 0.99$$

Taking log of both sides

$$-20p \log 2.72 \leq \log 0.99$$

$$-(20 \times 0.4346)p \geq -1.9956$$

No. of mistakes per page	:	0	1	2	3
Frequency	:	211	90	19	5

$$-8.692p \leq -0.0044 \text{ or } p \geq \frac{0.0044}{8.692} = 0.00051.$$

- (ii) In this case r is a Poisson variate which denotes the number of mistakes per day. Since the typist has to type $20 \times 200 = 4000$ words per day, the mean number of mistakes $= 4000p$.

It is given that there is no mistake on 90% of the days, i.e.,

$$P(r = 0) = 0.90 \text{ or } e^{-4000p} = 0.90$$

Taking log of both sides, we have

$$-4000p \log 2.72 = \log 0.90 \text{ or } -4000 \times 0.4346p = -1.9542 = -0.0458$$

$$\therefore p = \frac{0.0458}{4000 \times 0.4346} = 0.000026.$$

Example 20: A manufacturer of pins knows that on an average 5% of his product is defective. He sells pins in boxes of 100 and guarantees that not more than 4 pins will be defective. What is the probability that the box will meet the guaranteed quality?

Solution: The number of defective pins in a box is a Poisson variate with mean equal to 5. A box will meet the guaranteed quality if $r \leq 4$. Thus, the required probability is given by

$$P(r \leq 4) = e^{-5} \sum_{r=0}^4 \frac{5^r}{r!} = e^{-5} \left[1 + 5 + \frac{25}{2} + \frac{125}{6} + \frac{625}{24} \right] = 0.00678 \times \frac{1569}{24} = 0.44324.$$

Poisson Approximation to Binomial

When n , the number of trials become large, the computation of probabilities by using the binomial probability mass function becomes a cumbersome task. Usually, when $n \geq 20$ and $p \leq 0.05$, Poisson distribution can be used as an approximation to binomial with parameter $m = np$.

Example 21: Find the probability of 4 successes in 30 trials by using (i) binomial distribution and (ii) Poisson distribution. The probability of success in each trial is given to be 0.02.

Solution:

(i) Here $n = 30$ and $p = 0.02$

$$\therefore P(r=4) = {}^{30}C_4 (0.02)^4 (0.98)^{26} = 27405 \times 0.00000016 \times 0.59 = 0.00259.$$

(ii) Here $m = np = 30 \times 0.02 = 0.6$

$$\therefore P(r=4) = \frac{e^{-0.6} (0.6)^4}{4!} = \frac{0.5488 \times 0.1296}{24} = 0.00296.$$

Fitting of a Poisson Distribution

To fit a Poisson distribution to a given frequency distribution, we first compute its mean m . Then the probabilities of various values of the random variable r are computed by using the probability mass function $P(r) = \frac{e^{-m} \cdot m^r}{r!}$. These probabilities are then multiplied by N , the total frequency, to get expected frequencies.

Example 22: The following mistakes per page were observed in a book :

No. of mistakes per page:	0	1	2	3
Frequency	211	90	19	5

Fit a Poisson distribution to find the theoretical frequencies.

Solution: The mean of the given frequency distribution is

$$m = \frac{0 \times 211 + 1 \times 90 + 2 \times 19 + 3 \times 5}{211 + 90 + 19 + 5} = \frac{143}{325} = 0.44$$

Calculation of theoretical (or expected) frequencies

We can write $P(r) = \frac{e^{-0.44} (0.44)^r}{r!}$. Substituting $r = 0, 1, 2$ and 3 , we get the probabilities for various values of r , as shown in the following table:

r	$P(r)$	$N \times P(r)$	Expected Frequencies Approximated to the nearest integer
0	0.6440	209.30	210
1	0.2834	92.10	92
2	0.0623	20.25	20
3	0.0091	2.96	3
<i>Total</i>			325

Features of Poisson Distribution

- (i) It is discrete probability distribution.
- (ii) It has only one parameter m .

- (iii) The range of the random variable is $0 \leq r < \infty$.
- (iv) The Poisson distribution is a positively skewed distribution. The skewness decreases as m increases.

Uses of Poisson Distribution

- (i) This distribution is applicable to situations where the number of trials is large and the probability of a success in a trial is very small.
- (ii) It serves as a reasonably good approximation to binomial distribution when $n \geq 20$ and $p \leq 0.05$.

Exercise with Hints

1. If 2% of the electric bulbs manufactured by a company are defective, find the probability that in a sample of 200 bulbs (i) less than 2 bulbs are defective, (ii) more than 3 bulbs are defective. (Given $e^{-4} = 0.0183$).

Hint: $m = \frac{2}{100} \times 200 = 4$.

2. If r is a Poisson variate such that $P(r) = P(r + 1)$, what are the mean and standard deviation of r ?

Hint: Find m by using the given condition.

3. The number of arrivals of telephone calls at a switch board follows a Poisson process at an average rate of 8 calls per 10 minutes. The operator leaves for a 5 minutes tea break. Find the probability that (a) at the most two calls go unanswered and (b) 3 calls go unanswered, while the operator is away.

Hint: $m = 4$.

4. What probability model is appropriate to describe a situation where 100 misprints are distributed randomly throughout the 100 pages of a book? For this model, what is the probability that a page observed at random will contain (i) no misprint, (ii) at the most two misprints, (iii) at least three misprints?

Hint: The average number of misprint per page is unity.

5. If the probability of getting a defective transistor in a consignment is 0.01, find the mean and standard deviation of the number of defective transistors in a large consignment of 900 transistors. What is the probability that there is at the most one defective transistor in the consignment?

Hint: The average number of transistors in a consignment is 900×0.01 .

6. In a certain factory turning out blades, there is a small chance 1/500 for any one blade to be defective. The blades are supplied in packets of 10. Use Poisson distribution to compute the approximate number of packets containing no defective, one defective, two defective, three defective blades respectively in a consignment of 10,000 packets.

Hint: The random variable is the number of defective blades in a packet of 10 blades.

7. A manufacturer knows that 0.3% of items produced in his factory are defective. If the items are supplied in boxes, each containing 250 items, what is the probability that a box contains (i) no defective, (ii) at the most two defective items?

Hint: $m = \frac{0.3}{100} \times 250 = 0.75$.

8. A random variable r follows Poisson distribution, where $P(r = 2) = P(r = 3)$. Find (i) $P(r = 0)$, (ii) $P(1 \leq r \leq 3)$.

Hint: $P(1 \leq r \leq 3) = P(r = 1) + P(r = 2) + P(r = 3)$.

9. If X is a Poisson variate such that $P(X = 2) = 9P(X = 4) + 90P(X = 6)$, find the mean and variance of X .

Hint: mean = Variance.

10. Lots of 400 wall-clocks are purchased by a retailer. The retailer inspects sample of 20 clocks from each lot and returns the lot to the supplier if there are more than two defectives in the sample. Suppose a lot containing 30 defective clocks is received by the retailer, what is the probability that it will be returned to the supplier?

Hint: $n = 20$ and $p = 30/400$.

11. An industrial area has power breakdown once in 15 days, on the average. Assuming that the number of breakdowns follow a Poisson process, what is the probability of (i) no power breakdown in the next six days, (ii) more than one power breakdown in the next six days?

Hint: The random variable is the number of power breakdowns in six days.

12. After correcting the proofs of first 50 pages or so of a book, it is found that on the average there are 3 errors per 5 pages. Use Poisson probabilities and estimate the number of pages with 0, 1, 2, 3, errors in the whole book of 1,000 pages. [Given that $e^{-0.6} = 0.5488$].

Hint: Take random variable as the number of errors per page.

13. Between 2 and 4 p.m., the number of phone calls coming into the switch board of a company is 300. Find the probability that during one particular minute there will be (i) no phone call at all, (ii) exactly 3 calls, (iii) at least 7 calls. [Given $e^{-2} = 0.13534$ and $e^{-0.5} = 0.60650$].

Hint: Random variable is the number of calls per minute.

14. It is known that 0.5% of ball pen refills produced by a factory are defective. These refills are dispatched in packagings of equal numbers. Using Poisson distribution determine the number of refills in a packing to be sure that at least 95% of them contain no defective refills.

Hint: Let n be the number of refills in a package, then $m = 0.005n$.

15. Records show that the probability is 0.00002 that a car will have a flat tyre while driving over a certain bridge. Find the probability that out of 20,000 cars driven over the bridge, not more than one will have a flat tyre.

Hint: The random variable is number of cars driven over the bridge having flat tyre.

16. A radioactive source emits on the average 2.5 particles per second. Calculate the probability that two or more particles will be emitted in an interval of 4 seconds.

Hint: $m = 2.5 \times 4$.

17. The number of accidents in a year attributed to taxi drivers in a city follows Poisson distribution with mean 3. Out of 1,000 taxi drivers, find approximately the number of drivers with (i) no accident in a year, (ii) more than 3 accidents in a year. [Given $e^{-1} = 0.3679$, $e^{-2} = 0.1353$, $e^{-3} = 0.0498$].

Hint: Number of drivers = probability $\times 1000$.

18. A big industrial plant has to be shut down for repairs on an average of 3 times in a month. When more than 5 shut downs occur for repairs in a month, the production schedule cannot be attained. Find the probability that production schedule cannot be attained in a given month, assuming that the number of shut downs are a Poisson variate.

Hint: Find $P(r \geq 5)$.

19. A manager receives an average of 12 telephone calls per 8-hour day. Assuming that the number of telephone calls received by him follow a Poisson variate, what is the probability that he will not be interrupted by a call during a meeting lasting 2 hours?

Hint: Take $m = 3$.

20. Assuming that the probability of a fatal accident in a factory during a year is $1/1200$, calculate the probability that in a factory employing 300 workers, there will be at least two fatal accidents in a year. [Given $e^{-0.25} = 0.7788$].

Hint: The average number of accidents per year in the factory = 0.25.

21. If 2% of electric bulbs manufactured by a certain company are defective, find the probability that in a sample of 200 bulbs (i) less than 2 bulbs are defective (ii) more than 3 bulbs are defective. [Given $e^{-4} = 0.0183$].

Hint: $m = 4$.

22. If for a Poisson variate X , $P(X = 1) = P(X = 2)$, find $P(X = 1 \text{ or } 2)$. Also find its mean and standard deviation.

Hint: Find m from the given condition.

23. If 5% of the families in Calcutta do not use gas as a fuel, what will be the probability of selecting 10 families in a random sample of 100 families who do not use gas as a fuel? You may assume Poisson distribution. [Given $e^{-5} = 0.0067$].

Hint: $m = 5$, find $P(r = 10)$.

24. The probability that a Poisson variate X takes a positive value is $1 - e^{-1.5}$. Find the variance and also the probability that X lies between -1.5 and 1.5 .

Hint: $1 - e^{-1.5} = P(r > 0)$. Find $P(-1.5 < X < 1.5) = P(X = 0) + P(X = 1)$.

25. 250 passengers have made reservations for a flight from Delhi to Mumbai. If the probability that a passenger, who has reservation, will not turn up is 0.016, find the probability that at the most 3 passengers will not turn up.

Hint: The number of passengers who do not turn up is a Poisson variate.

11.9 EXPONENTIAL DISTRIBUTION

The random variable in case of Poisson distribution is of the type ; the number of arrivals of customers per unit of time or the number of defects per unit length of cloth, etc. Alternatively, it is possible to define a random variable, in the context of *Poisson Process*, as the length of time between the arrivals of two consecutive customers or the length of cloth between two consecutive defects, etc. The probability distribution of such a random variable is termed as *Exponential Distribution*.

Since the length of time or distance is a continuous random variable, therefore exponential distribution is a continuous probability distribution.

Probability Density Function

Let t be a random variable which denotes the length of time or distance between the occurrence of two consecutive events or the occurrence of the first event and m be the average number of times the event occurs per unit of time or length. Further, let A be the event that the time of occurrence between two consecutive events or the occurrence of the first event is less than or equal to t and $f(t)$ and $F(t)$ denote the probability density function and the distribution (or cumulative density) function of t respectively.

We can write $P(A) + P(\bar{A}) = 1$ or $F(t) + P(\bar{A}) = 1$. Note that, by definition, $F(t) = P(A)$

Further, $P(\bar{A})$ is the probability that the length of time between the occurrence of two consecutive events or the occurrence of first event is greater than t . This is also equal to

the probability that no event occurs in the time interval t . Since the mean number of occurrence of events in time t is mt , we have , by Poisson distribution,

$$P(\bar{A}) = P(r=0) = \frac{e^{-mt} (mt)^0}{0!} = e^{-mt}.$$

Thus, we get $F(t) + e^{-mt} = 1$

$$\text{or } P(0 \text{ to } t) = F(t) = 1 - e^{-mt}. \quad \dots (1)$$

To get the probability density function, we differentiate equation (1) with respect to t .

$$\begin{aligned} \text{Thus, } f(t) &= F'(t) = me^{-mt} \quad \text{when } t > 0 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

It can be verified that the total probability is equal to unity

$$\text{Total Probability} = \int_0^\infty m.e^{-mt} dt = \left| m \cdot \frac{e^{-mt}}{-m} \right|_0^\infty = \left| -e^{-mt} \right|_0^\infty = 0 + 1 = 1.$$

Mean of t

The mean of t is defined as its expected value, given by

$$E(t) = \int_0^\infty t.m.e^{-mt} dt = \frac{1}{m}, \text{ where } m \text{ denotes the average number of occurrence of events per unit of time or distance.}$$

Example 23: A telephone operator attends on an average 150 telephone calls per hour. Assuming that the distribution of time between consecutive calls follows an exponential distribution, find the probability that (i) the time between two consecutive calls is less than 2 minutes, (ii) the next call will be received only after 3 minutes.

Solution: Here m = the average number of calls per minute = $\frac{150}{60} = 2.5$.

$$(i) \quad P(t \leq 2) = \int_0^2 2.5e^{-2.5t} dt = F(2)$$

$$\text{We know that } F(t) = 1 - e^{-mt}, \therefore F(2) = 1 - e^{-2.5 \times 2} = 0.9933$$

$$\begin{aligned} (ii) \quad P(t > 3) &= 1 - P(t \leq 3) = 1 - F(3) \\ &= 1 - [1 - e^{-2.5 \times 3}] = 0.0006 \end{aligned}$$

Example 24: The average number of accidents in an industry during a year is estimated to be 5. If the distribution of time between two consecutive accidents is known to be exponential, find the probability that there will be no accidents during the next two months.

Solution: Here m denotes the average number of accidents per month = $\frac{5}{12}$.

$$\therefore P(t > 2) = 1 - F(2) = e^{-\frac{5}{12} \times 2} = e^{-0.833} = 0.4347.$$

Example 25: The distribution of life, in hours, of a bulb is known to be exponential with mean life of 600 hours. What is the probability that (i) it will not last more than 500 hours, (ii) it will last more than 700 hours?

Solution: Since the random variable denote hours, therefore $m = \frac{1}{600}$

$$(i) \quad P(t \leq 500) = F(500) = 1 - e^{-\frac{1}{600} \times 500} = 1 - e^{-0.833} = 0.5653.$$

$$(ii) \quad P(t > 700) = 1 - F(700) = e^{-\frac{700}{600}} = e^{-1.1667} = 0.3114.$$

11.10 UNIFORM DISTRIBUTION (CONTINUOUS VARIABLE)

A continuous random variable X is said to be uniformly distributed in a close interval

(α, β) with probability density function $p(X)$

$$\text{if } p(X) = \frac{1}{\beta - \alpha} \text{ for } \alpha \leq X \leq \beta \text{ and } = 0$$

Otherwise The uniform distribution is alternatively known as rectangular distribution.

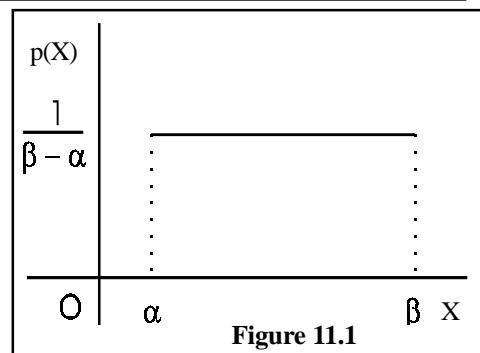


Figure 11.1

The diagram of the probability density function is shown in the figure 19.1.

Note that the total area under the curve is unity, i.e. ,

$$\int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} dX = \frac{1}{\beta - \alpha} |X|_{\alpha}^{\beta} = 1$$

$$\text{Further, } E(X) = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} X \cdot dX = \frac{1}{\beta - \alpha} \left| \frac{X^2}{2} \right|_{\alpha}^{\beta} = \frac{\alpha + \beta}{2}$$

$$E(X^2) = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} X^2 \cdot dX = \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} = \frac{1}{3} (\beta^2 + \alpha\beta + \alpha^2)$$

$$\therefore \text{Var}(X) = \frac{1}{3} (\beta^2 + \alpha\beta + \alpha^2) - \frac{(\alpha + \beta)^2}{4} = \frac{(\beta - \alpha)^2}{12}$$

Example 26: The buses on a certain route run after every 20 minutes. If a person arrives at the bus stop at random, what is the probability that

- (a) he has to wait between 5 to 15 minutes,
- (b) he gets a bus within 10 minutes,
- (c) he has to wait at least 15 minutes.

Solution: Let the random variable X denote the waiting time, which follows a uniform distribution with p.d.f.

$$f(X) = \frac{1}{20} \quad \text{for } 0 \leq X \leq 20$$

$$(a) P(5 \leq X \leq 15) = \frac{1}{20} \int_5^{15} dX = \frac{1}{20} (15 - 5) = \frac{1}{2}$$

$$(b) P(0 \leq X \leq 10) = \frac{1}{20} \times 10 = \frac{1}{2}$$

$$(c) P(15 \leq X \leq 20) = \frac{20 - 15}{20} = \frac{1}{4}.$$

11.11 NORMAL DISTRIBUTION

The normal probability distribution occupies a place of central importance in *Modern Statistical Theory*. This distribution was first observed as the *normal law of errors* by the statisticians of the eighteenth century. They found that each observation X involves an error term which is affected by a large number of small but independent chance factors. This implies that an observed value of X is the sum of its true value and the net effect of a large number of independent errors which may be positive or negative each with equal probability. The observed distribution of such a random variable was found to be in close conformity with a continuous curve, which was termed as the *normal curve of errors* or simply the *normal curve*.

Since Gauss used this curve to describe the theory of accidental errors of measurements involved in the calculation of orbits of heavenly bodies, it is also called as Gaussian curve.

The Conditions of Normality

In order that the distribution of a random variable X is normal, the factors affecting its observations must satisfy the following conditions :

- (i) **A large number of chance factors:** The factors, affecting the observations of a random variable, should be numerous and equally probable so that the occurrence or non-occurrence of any one of them is not predictable.
- (ii) **Condition of homogeneity:** The factors must be similar over the relevant population although, their incidence may vary from observation to observation.
- (iii) **Condition of independence:** The factors, affecting observations, must act independently of each other.
- (iv) **Condition of symmetry:** Various factors operate in such a way that the deviations of observations above and below mean are balanced with regard to their magnitude as well as their number.

Random variables observed in many phenomena related to economics, business and other social as well as physical sciences are often found to be distributed normally. For example, observations relating to the life of an electrical component, weight of packages, height of persons, income of the inhabitants of certain area, diameter of wire, etc., are affected by a large number of factors and hence, tend to follow a pattern that is very similar to the normal curve. In addition to this, when the number of observations become large, a number of probability distributions like Binomial, Poisson, etc., can also be approximated by this distribution.

Probability Density Function

If X is a continuous random variable, distributed normally with mean m and standard deviation σ , then its p.d.f. is given by

$$p(X) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad \text{where } -\infty < X < \infty.$$

Here π and σ are absolute constants with values 3.14159.... and 2.71828.... respectively.

It may be noted here that this distribution is completely known if the values of mean m and standard deviation s are known. Thus, the distribution has two parameters, viz. mean and standard deviation.

Shape of Normal Probability Curve

For given values of the parameters, m and s , the shape of the curve corresponding to normal probability density function $p(X)$ is as shown in Figure. 11.2

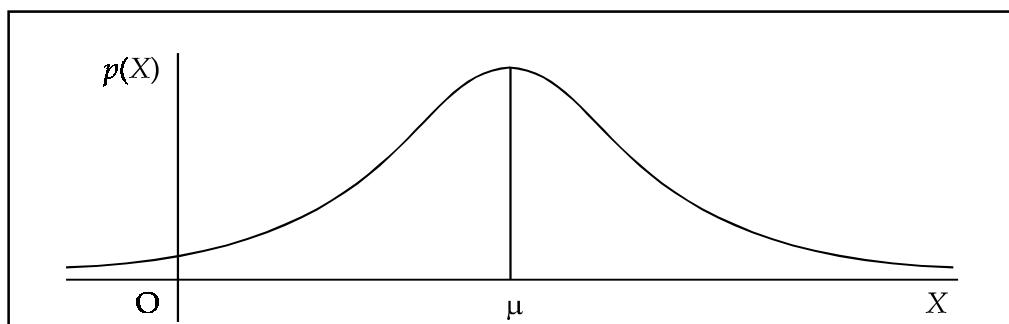


Figure 11.2

It should be noted here that although we seldom encounter variables that have a range from $-\infty$ to ∞ , as shown by the normal curve, nevertheless the curves generated by the relative frequency histograms of various variables closely resembles the shape of normal curve.

Properties of Normal Probability Curve

A normal probability curve or normal curve has the following properties :

1. It is a bell shaped symmetrical curve about the ordinate at $X = \mu$. The ordinate is maximum at $X = \mu$.
2. It is unimodal curve and its tails extend infinitely in both directions, i.e., the curve is asymptotic to X axis in both directions.
3. All the three measures of central tendency coincide, i.e.,

$$\text{mean} = \text{median} = \text{mode}$$

4. The total area under the curve gives the total probability of the random variable taking values between $-\infty$ to ∞ . Mathematically, it can be shown that

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} p(X) dX = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} dX = 1.$$

5. Since median = m , the ordinate at $X = \mu$ divides the area under the normal curve into two equal parts, i.e.,

$$\int_{-\infty}^{\mu} p(X) dX = \int_{\mu}^{\infty} p(X) dX = 0.5$$

6. The value of $p(X)$ is always non-negative for all values of X , i.e., the whole curve lies above X axis.
7. The points of inflexion (the point at which curvature changes) of the curve are at $X = \mu \pm \sigma$.
8. The quartiles are equidistant from median, i.e., $M_d - Q_1 = Q_3 - M_d$, by virtue of symmetry. Also $Q_1 = \mu - 0.6745\sigma$, $Q_3 = \mu + 0.6745\sigma$, quartile deviation = 0.6745σ and mean deviation = 0.8σ , approximately.
9. Since the distribution is symmetrical, all odd ordered central moments are zero.
10. The successive even ordered central moments are related according to the following recurrence formula

$$\mu_{2n} = (2n - 1)\sigma^2 \mu_{2n-2} \quad \text{for } n = 1, 2, 3, \dots$$

11. The value of moment coefficient of skewness β_1 is zero.
12. The coefficient of kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{\sigma^4} = 3$.

Note that the above expression makes use of property 10.

13. Additive or reproductive property

If X_1, X_2, \dots, X_n are n independent normal variates with means $\mu_1, \mu_2, \dots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$, respectively, then their linear combination $a_1X_1 + a_2X_2 + \dots + a_nX_n$ is also a normal variate with mean $\sum_{i=1}^n a_i\mu_i$ and variance $\sum_{i=1}^n a_i^2\sigma_i^2$.

In particular, if $a_1 = a_2 = \dots = a_n = 1$, we have $\sum X_i$ is a normal variate with mean $\sum \mu_i$ and variance $\sum \sigma_i^2$. Thus the sum of independent normal variates is also a normal variate.

14. **Area property:** The area under the normal curve is distributed by its standard deviation in the following manner :

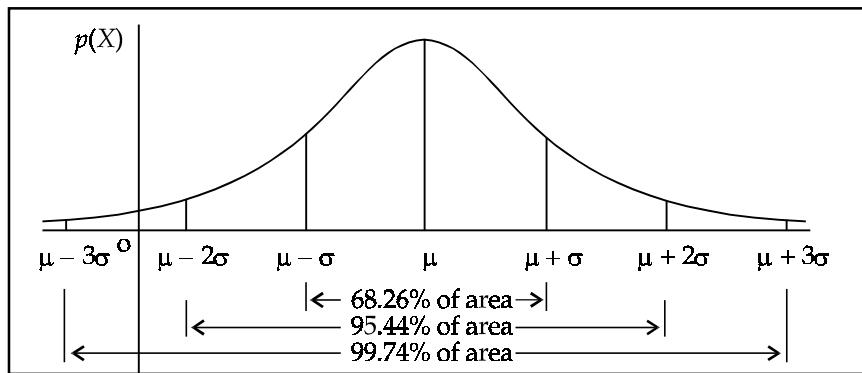


Figure 11.3

- (i) The area between the ordinates at $\mu - \sigma$ and $\mu + \sigma$ is 0.6826. This implies that for a normal distribution about 68% of the observations will lie between $\mu - \sigma$ and $\mu + \sigma$.
- (ii) The area between the ordinates at $\mu - 2\sigma$ and $\mu + 2\sigma$ is 0.9544. This implies that for a normal distribution about 95% of the observations will lie between $\mu - 2\sigma$ and $\mu + 2\sigma$.
- (iii) The area between the ordinates at $\mu - 3\sigma$ and $\mu + 3\sigma$ is 0.9974. This implies that for a normal distribution about 99% of the observations will lie between $\mu - 3\sigma$ and $\mu + 3\sigma$. This result shows that, practically, the range of the distribution is 6σ although, theoretically, the range is from $-\infty$ to ∞ .

Probability of Normal Variate in an Interval

Let X be a normal variate distributed with mean m and standard deviation s , also written in abbreviated form as $X \sim N(\mu, \sigma)$. The probability of X lying in the interval (X_1, X_2) is given by

$$P(X_1 \leq X \leq X_2) = \int_{X_1}^{X_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} dX$$

In terms of figure, this probability is equal to the area under the normal curve between the ordinates at $X = X_1$ and $X = X_2$ respectively.

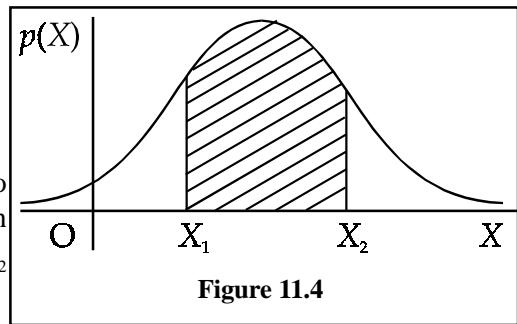


Figure 11.4

Note: It may be recalled that the probability that a continuous random variable takes a particular value is defined to be zero even though the event is not impossible.

It is obvious from the above that, to find $P(X_1 \leq X \leq X_2)$, we have to evaluate an integral which might be cumbersome and time consuming task. Fortunately, an alternative procedure is available for performing this task. To devise this procedure, we define a

$$\text{new variable } z = \frac{X - \mu}{\sigma}.$$

$$\text{We note that } E(z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}[E(X) - \mu] = 0$$

$$\text{and } Var(z) = Var\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} Var(X - \mu) = \frac{1}{\sigma^2} Var(X) = 1.$$

Further, from the reproductive property, it follows that the distribution of z is also normal.

Thus, we conclude that if X is a normal variate with mean m and standard deviation s , then $z = \frac{X - \mu}{\sigma}$ is a normal variate with mean zero and standard deviation unity.

Since the parameters of the distribution of z are fixed, it is a known distribution and is termed as *standard normal distribution* (s.n.d.). Further, z is termed as a *standard normal variate* (s.n.v.).

It is obvious from the above that the distribution of any normal variate X can always be transformed into the distribution of standard normal variate z . This fact can be utilised to evaluate the integral given above.

$$\text{We can write } P(X_1 \leq X \leq X_2) = P\left[\left(\frac{X_1 - \mu}{\sigma}\right) \leq \left(\frac{X - \mu}{\sigma}\right) \leq \left(\frac{X_2 - \mu}{\sigma}\right)\right]$$

$$= P(z_1 \leq z \leq z_2), \text{ where } z_1 = \frac{X_1 - \mu}{\sigma} \text{ and}$$

$$z_2 = \frac{X_2 - \mu}{\sigma}$$

In terms of figure, this probability is equal to the area under the standard normal curve between the ordinates at $z = z_1$ and $z = z_2$.

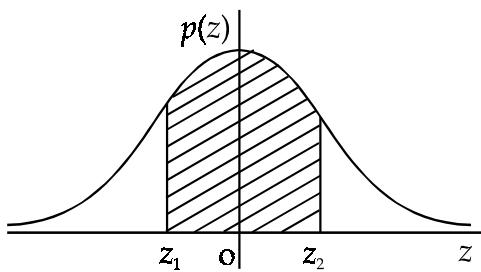


Figure 11.5

Since the distribution of z is fixed, the probabilities of z lying in various intervals are tabulated. These tables can be used to write down the desired probability.

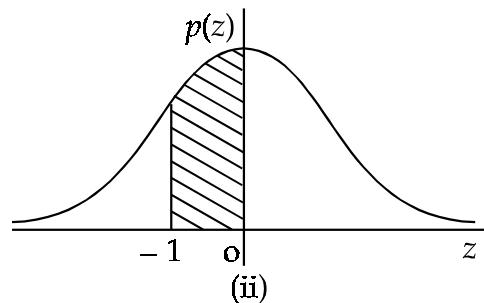
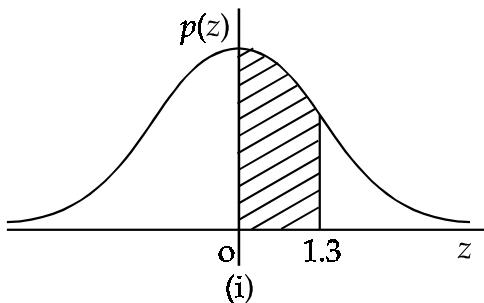
Example 27: Using the table of areas under the standard normal curve, find the following probabilities :

- (i) $P(0 \leq z \leq 1.3)$
- (ii) $P(-1 \leq z \leq 0)$
- (iii) $P(-1 \leq z \leq 12)$
- (iv) $P(z \geq 1.54)$
- (v) $P(|z| > 2)$
- (vi) $P(|z| < 2)$

Solution: The required probability, in each question, is indicated by the shaded area of the corresponding figure.

(i) From the table, we can write $P(0 \leq z \leq 1.3) = 0.4032$.

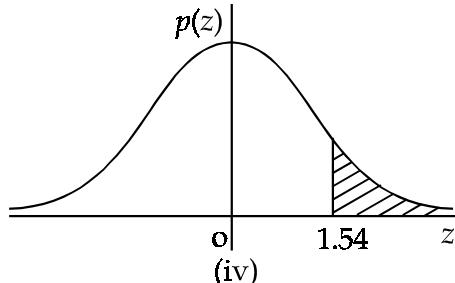
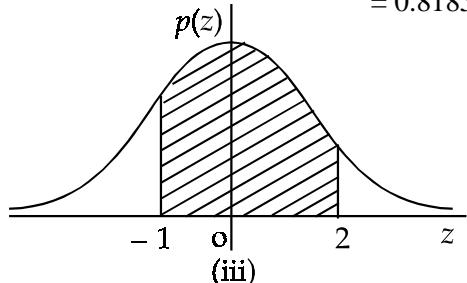
(ii) We can write $P(-1 \leq z \leq 0) = P(0 \leq z \leq 1)$, because the distribution is symmetrical.



From the table, we can write $P(-1 \leq z \leq 0) = P(0 \leq z \leq 1) = 0.3413$.

(iii) We can write $P(-1 \leq z \leq 2) = P(-1 \leq z \leq 0) + P(0 \leq z \leq 2)$

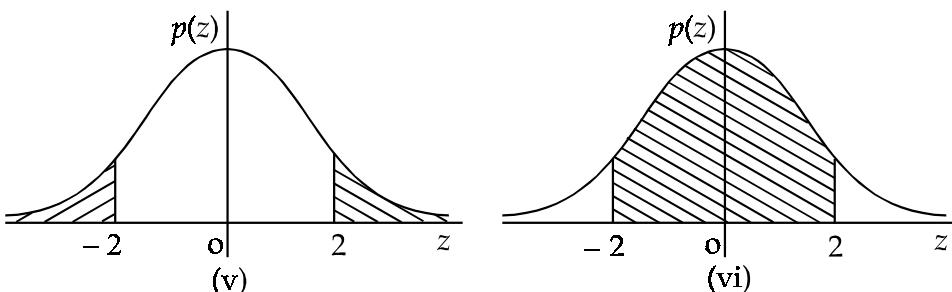
$$\begin{aligned} &= P(0 \leq z \leq 1) + P(0 \leq z \leq 2) = 0.3413 + 0.4772 \\ &= 0.8185. \end{aligned}$$



(iv) We can write

$$P(z \geq 1.54) = 0.5000 - P(0 \leq z \leq 1.54) = 0.5000 - 0.4382 = 0.0618.$$

(v) $P(|z| > 2) = P(z > 2) + P(z < -2) = 2P(z > 2) = 2[0.5000 - P(0 \leq z \leq 2)]$
 $= 1 - 2P(0 \leq z \leq 2) = 1 - 2 \times 0.4772 = 0.0456.$



(vi) $P(|z| < 2) = P(-2 \leq z \leq 0) + P(0 \leq z \leq 2) = 2P(0 \leq z \leq 2) = 2 \times 0.4772 = 0.9544.$

Example 28: Determine the value or values of z in each of the following situations:

- (a) Area between 0 and z is 0.4495.
- (b) Area between $-\infty$ to z is 0.1401.
- (c) Area between $-\infty$ to z is 0.6103.
- (d) Area between -1.65 and z is 0.0173.
- (e) Area between -0.5 and z is 0.5376.

Solution:

- (a) On locating the value of z corresponding to an entry of area 0.4495 in the table of areas under the normal curve, we have $z = 1.64$. We note that the same situation may correspond to a negative value of z. Thus, z can be 1.64 or -1.64.
- (b) Since the area between $-\infty$ to $z < 0.5$, z will be negative. Further, the area between z and 0 = $0.5000 - 0.1401 = 0.3599$. On locating the value of z corresponding to this entry in the table, we get $z = -1.08$.
- (c) Since the area between $-\infty$ to $z > 0.5000$, z will be positive. Further, the area between 0 to z = $0.6103 - 0.5000 = 0.1103$. On locating the value of z corresponding to this entry in the table, we get $z = 0.28$.
- (d) Since the area between -1.65 and z < the area between -1.65 and 0 (which, from table, is 0.4505), z is negative. Further z can be to the right or to the left of the value -1.65 . Thus, when z lies to the right of -1.65 , its value, corresponds to an area $(0.4505 - 0.0173) = 0.4332$, is given by $z = -1.5$ (from table). Further, when z lies to the left of -1.65 , its value, corresponds to an area $(0.4505 + 0.0173) = 0.4678$, is given by $z = -1.85$ (from table).
- (e) Since the area between -0.5 to z > area between -0.5 to 0 (which, from table, is 0.1915), z is positive. The value of z, located corresponding to an area $(0.5376 - 0.1915) = 0.3461$, is given by 1.02.

Example 29: If X is a random variate which is distributed normally with mean 60 and standard deviation 5, find the probabilities of the following events :

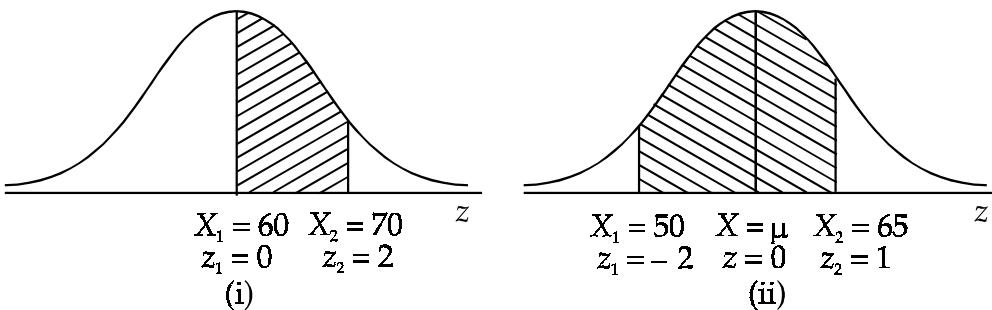
- (i) $60 \leq X \leq 70$, (ii) $50 \leq X \leq 65$, (iii) $X > 45$, (iv) $X \leq 50$.

Solution: It is given that $\mu = 60$ and $\sigma = 5$

- (i) Given $X_1 = 60$ and $X_2 = 70$, we can write

$$z_1 = \frac{X_1 - \mu}{\sigma} = \frac{60 - 60}{5} = 0 \text{ and } z_2 = \frac{X_2 - \mu}{\sigma} = \frac{70 - 60}{5} = 2.$$

$$\therefore P(60 \leq X \leq 70) = P(0 \leq z \leq 2) = 0.4772 \text{ (from table).}$$

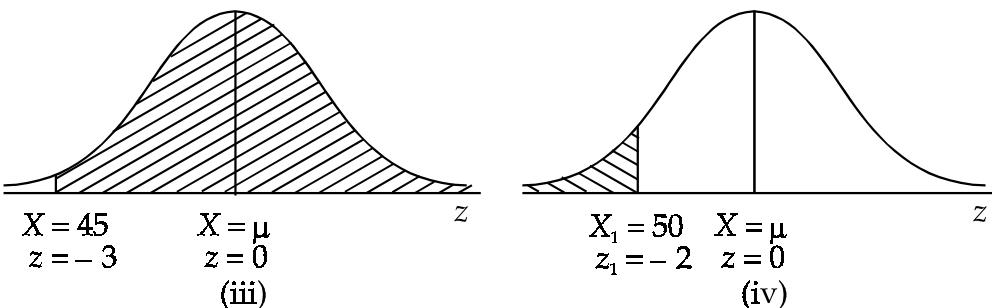


(ii) Here $X_1 = 50$ and $X_2 = 65$, therefore, we can write

$$z_1 = \frac{50 - 60}{5} = -2 \text{ and } z_2 = \frac{65 - 60}{5} = 1.$$

$$\begin{aligned} \text{Hence } P(50 \leq X \leq 65) &= P(-2 \leq z \leq 1) = P(0 \leq z \leq 2) + P(0 \leq z \leq 1) \\ &= 0.4772 + 0.3413 = 0.8185 \end{aligned}$$

$$\begin{aligned} (\text{iii}) \quad P(X > 45) &= P\left(z \geq \frac{45 - 60}{5}\right) = P(z \geq -3) \\ &= P(-3 \leq z \leq 0) + P(0 \leq z \leq \infty) = P(0 \leq z \leq 3) + P(0 \leq z \leq \infty) \\ &= 0.4987 + 0.5000 = 0.9987 \end{aligned}$$



$$\begin{aligned} (\text{iv}) \quad P(X \leq 50) &= P\left(z \leq \frac{50 - 60}{5}\right) = P(z \leq -2) \\ &= 0.5000 - P(-2 \leq z \leq 0) = 0.5000 - P(0 \leq z \leq 2) \\ &= 0.5000 - 0.4772 = 0.0228 \end{aligned}$$

Example 30: The average monthly sales of 5,000 firms are normally distributed with mean Rs 36,000 and standard deviation Rs 10,000. Find :

- (i) The number of firms with sales of over Rs 40,000.
- (ii) The percentage of firms with sales between Rs 38,500 and Rs 41,000.
- (iii) The number of firms with sales between Rs 30,000 and Rs 40,000.

Solution: Let X be the normal variate which represents the monthly sales of a firm. Thus $X \sim N(36,000, 10,000)$.

$$\begin{aligned} (\text{i}) \quad P(X > 40000) &= P\left(z > \frac{40000 - 36000}{10000}\right) = P(z > 0.4) \\ &= 0.5000 - P(0 \leq z \leq 0.4) = 0.5000 - 0.1554 = 0.3446. \end{aligned}$$

Thus, the number of firms having sales over Rs 40,000

$$= 0.3446 \times 5000 = 1723$$

$$\begin{aligned}
 \text{(ii)} \quad P(38500 \leq X \leq 41000) &= P\left(\frac{38500 - 36000}{10000} \leq z \leq \frac{41000 - 36000}{10000}\right) \\
 &= P(0.25 \leq z \leq 0.5) = P(0 \leq z \leq 0.5) - P(0 \leq z \leq 0.25) \\
 &= 0.1915 - 0.0987 = 0.0987.
 \end{aligned}$$

Thus, the required percentage of firms = $0.0987 \times 100 = 9.87\%$.

$$\begin{aligned}
 \text{(iii)} \quad P(30000 \leq X \leq 40000) &= P\left(\frac{30000 - 36000}{10000} \leq z \leq \frac{40000 - 36000}{10000}\right) \\
 &= P(-0.6 \leq z \leq 0.4) = P(0 \leq z \leq 0.6) + P(0 \leq z \leq 0.4) \\
 &= 0.2258 + 0.1554 = 0.3812.
 \end{aligned}$$

Thus, the required number of firms = $0.3812 \times 5000 = 1906$

Example 31: In a large institution, 2.28% of employees have income below Rs 4,500 and 15.87% of employees have income above Rs. 7,500 per month. Assuming the distribution of income to be normal, find its mean and standard deviation.

Solution: Let the mean and standard deviation of the given distribution be μ and σ respectively.

It is given that $P(X < 4500) = 0.0228$ or $P\left(z < \frac{4500 - \mu}{\sigma}\right) = 0.0228$

On locating the value of z corresponding to an area 0.4772 (0.5000 - 0.0228), we can

$$\text{write } \frac{4500 - \mu}{\sigma} = -2 \text{ or } 4500 - \mu = -2\sigma \quad \dots (1)$$

Similarly, it is also given that

$$P(X > 7500) = 0.1587 \text{ or } P\left(z > \frac{7500 - \mu}{\sigma}\right) = 0.1587$$

Locating the value of z corresponding to an area 0.3413 (0.5000 - 0.1587), we can write

$$\frac{7500 - \mu}{\sigma} = 1 \text{ or } 7500 - \mu = \sigma \quad \dots (2)$$

Solving (1) and (2) simultaneously, we get

$$\mu = \text{Rs } 6,500 \text{ and } \sigma = \text{Rs } 1,000.$$

Example 32: Marks in an examination are approximately normally distributed with mean 75 and standard deviation 5. If the top 5% of the students get grade A and the bottom 25% get grade F, what mark is the lowest A and what mark is the highest F?

Solution: Let A be the lowest mark in grade A and F be the highest mark in grade F. From the given information, we can write

$$P(X \geq A) = 0.05 \text{ or } P\left(z \geq \frac{A - 75}{5}\right) = 0.05$$

On locating the value of z corresponding to an area 0.4500 (0.5000 - 0.0500), we can

$$\text{write } \frac{A - 75}{5} = 1.645 \Rightarrow A = 83.225$$

Further, it is given that

$$P(X \leq F) = 0.25 \text{ or } P\left(z \leq \frac{F - 75}{5}\right) = 0.25$$

On locating the value of z corresponding to an area 0.2500 ($0.5000 - 0.2500$), we can write $\frac{F-75}{5} = -0.675 \Rightarrow F = 71.625$

Example 33: The mean inside diameter of a sample of 200 washers produced by a machine is 5.02 mm and the standard deviation is 0.05 mm. The purpose for which these washers are intended allows a maximum tolerance in the diameter of 4.96 to 5.08 mm, otherwise the washers are considered as defective. Determine the percentage of defective washers produced by the machine on the assumption that diameters are normally distributed.

Solution: Let X denote the diameter of the washer. Thus, $X \sim N(5.02, 0.05)$.

The probability that a washer is defective $= 1 - P(4.96 \leq X \leq 5.08)$

$$\begin{aligned} &= 1 - P\left[\left(\frac{4.96 - 5.02}{0.05}\right) \leq z \leq \left(\frac{5.08 - 5.02}{0.05}\right)\right] \\ &= 1 - P(-1.2 \leq z \leq 1.2) = 1 - 2P(0 \leq z \leq 1.2) = 1 - 2 \times 0.3849 = 0.2302 \end{aligned}$$

Thus, the percentage of defective washers $= 23.02$.

Example 34: The average number of units produced by a manufacturing concern per day is 355 with a standard deviation of 50. It makes a profit of Rs 1.50 per unit. Determine the percentage of days when its total profit per day is (i) between Rs 457.50 and Rs 645.00, (ii) greater than Rs 682.50 (assume the distribution to be normal). The area between $z = 0$ to $z = 1$ is 0.34134, the area between $z = 0$ to $z = 1.5$ is 0.43319 and the area between $z = 0$ to $z = 2$ is 0.47725, where z is a standard normal variate.

Solution: Let X denote the profit per day. The mean of X is $355 \times 1.50 = \text{Rs } 532.50$ and its S.D. is $50 \times 1.50 = \text{Rs } 75$. Thus, $X \sim N(532.50, 75)$.

(i) The probability of profit per day lying between Rs 457.50 and Rs 645.00

$$\begin{aligned} P(457.50 \leq X \leq 645.00) &= P\left(\frac{457.50 - 532.50}{75} \leq z \leq \frac{645.00 - 532.50}{75}\right) \\ &= P(-1 \leq z \leq 1.5) = P(0 \leq z \leq 1) + P(0 \leq z \leq 1.5) = 0.34134 + 0.43319 = 0.77453 \end{aligned}$$

Thus, the percentage of days $= 77.453$

$$\begin{aligned} \text{(ii)} \quad P(X \geq 682.50) &= P\left(z \geq \frac{682.50 - 532.50}{75}\right) = P(z \geq 2) \\ &= 0.5000 - P(0 \leq z \leq 2) = 0.5000 - 0.47725 = 0.02275 \end{aligned}$$

Thus, the percentage of days $= 2.275$

Example 35: The distribution of 1,000 examinees according to marks percentage is given below :

% Marks	less than 40	40 - 75	75 or more	Total
No. of examinees	430	420	150	1000

Assuming the marks percentage to follow a normal distribution, calculate the mean and standard deviation of marks. If not more than 300 examinees are to fail, what should be the passing marks?

Solution: Let X denote the percentage of marks and its mean and S.D. be μ and σ respectively. From the given table, we can write

$P(X < 40) = 0.43$ and $P(X \geq 75) = 0.15$, which can also be written as

$$P\left(z < \frac{40 - \mu}{\sigma}\right) = 0.43 \text{ and } P\left(z \geq \frac{75 - \mu}{\sigma}\right) = 0.15$$

The above equations respectively imply that

$$\frac{40 - \mu}{\sigma} = -0.175 \text{ or } 40 - \mu = -0.175\sigma \quad \dots (1)$$

$$\text{and } \frac{75 - \mu}{\sigma} = 1.04 \text{ or } 75 - \mu = 1.04\sigma \quad \dots (2)$$

Solving the above equations simultaneously, we get $\mu = 45.04$ and $\sigma = 28.81$.

Let X_1 be the percentage of marks required to pass the examination.

$$\text{Then we have } P(X < X_1) = 0.3 \text{ or } P\left(z < \frac{X_1 - 45.04}{28.81}\right) = 0.3$$

$$\therefore \frac{X_1 - 45.04}{28.81} = -0.525 \Rightarrow X_1 = 29.91 \text{ or } 30\% \text{ (approx.)}$$

Example 36: In a certain book, the frequency distribution of the number of words per page may be taken as approximately normal with mean 800 and standard deviation 50. If three pages are chosen at random, what is the probability that none of them has between 830 and 845 words each?

Solution: Let X be a normal variate which denotes the number of words per page. It is given that $X \sim N(800, 50)$.

The probability that a page, selected at random, does not have number of words between 830 and 845, is given by

$$\begin{aligned} 1 - P(830 < X < 845) &= 1 - P\left(\frac{830 - 800}{50} < z < \frac{845 - 800}{50}\right) \\ &= 1 - P(0.6 < z < 0.9) = 1 - P(0 < z < 0.9) + P(0 < z < 0.6) \\ &= 1 - 0.3159 + 0.2257 = 0.9098 \approx 0.91 \end{aligned}$$

Thus, the probability that none of the three pages, selected at random, have number of words lying between 830 and 845 = $(0.91)^3 = 0.7536$.

Example 37: At a petrol station, the mean quantity of petrol sold to a vehicle is 20 litres per day with a standard deviation of 10 litres. If on a particular day, 100 vehicles took 25 or more litres of petrol, estimate the total number of vehicles who took petrol from the station on that day. Assume that the quantity of petrol taken from the station by a vehicle is a normal variate.

Solution: Let X denote the quantity of petrol taken by a vehicle. It is given that $X \sim N(20, 10)$.

$$\begin{aligned} \therefore P(X \geq 25) &= P\left(z \geq \frac{25 - 20}{10}\right) = P(z \geq 0.5) \\ &= 0.5000 - P(0 \leq z \leq 0.5) = 0.5000 - 0.1915 = 0.3085 \end{aligned}$$

Let N be the total number of vehicles taking petrol on that day.

$$\therefore 0.3085 \times N = 100 \text{ or } N = 100/0.3085 = 324 \text{ (approx.)}$$

Normal Approximation to Binomial Distribution

Theoretical Probability
Distributions

Normal distribution can be used as an approximation to binomial distribution when n is large and neither p nor q is very small. If X denotes the number of successes with probability p of a success in each of the n trials, then X will be distributed approximately normally with mean np and standard deviation \sqrt{npq} .

$$\text{Further, } z = \frac{X - np}{\sqrt{npq}} \sim N(0,1).$$

It may be noted here that as X varies from 0 to n , the standard normal variate z would vary from $-\infty$ to ∞ because

$$\text{when } X = 0, \lim_{n \rightarrow \infty} \left(\frac{-np}{\sqrt{npq}} \right) \left(-\sqrt{\frac{np}{q}} \right) = -\infty$$

$$\text{and when } X = n, \lim_{n \rightarrow \infty} \left(\frac{n - np}{\sqrt{npq}} \right) = \lim_{n \rightarrow \infty} \left(\frac{nq}{\sqrt{npq}} \right) = \lim_{n \rightarrow \infty} \left(\sqrt{\frac{nq}{p}} \right) = \infty$$

Correction for Continuity

Since the number of successes is a discrete variable, to use normal approximation, we have make corrections for continuity. For example,

$P(X_1 \leq X \leq X_2)$ is to be corrected as $P\left(X_1 - \frac{1}{2} \leq X \leq X_2 + \frac{1}{2}\right)$, while using normal approximation to binomial since the gap between successive values of a binomial variate is unity. Similarly, $P(X_1 < X < X_2)$ is to be corrected as $P\left(X_1 + \frac{1}{2} \leq X \leq X_2 - \frac{1}{2}\right)$, since $X_1 < X$ does not include X_1 and $X < X_2$ does not include X_2 .

Note: The normal approximation to binomial probability mass function is good when $n \geq 50$ and neither p nor q is less than 0.1.

Example 38: An unbiased die is tossed 600 times. Use normal approximation to binomial to find the probability obtaining

- (i) more than 125 aces,
- (ii) number of aces between 80 and 110,
- (iii) exactly 150 aces.

Solution: Let X denote the number of successes, i.e., the number of aces.

$$\therefore \mu = np = 600 \times \frac{1}{6} = 100 \text{ and } \sigma = \sqrt{npq} = \sqrt{600 \times \frac{1}{6} \times \frac{5}{6}} = 9.1$$

- (i) To make correction for continuity, we can write

$$P(X > 125) = P(X > 125 + 0.5)$$

$$\text{Thus, } P(X \geq 125.5) = P\left(z \geq \frac{125.5 - 100}{9.1}\right) = P(z \geq 2.80)$$

$$= 0.5000 - P(0 \leq z \leq 2.80) = 0.5000 - 0.4974 = 0.0026.$$

- (ii) In a similar way, the probability of the number of aces between 80 and 110 is given by

$$\begin{aligned} P(79.5 \leq X \leq 110.5) &= P\left(\frac{79.5 - 100}{9.1} \leq z \leq \frac{110.5 - 100}{9.1}\right) \\ &= P(-2.25 \leq z \leq 1.15) = P(0 \leq z \leq 2.25) + P(0 \leq z \leq 1.15) \\ &= 0.4878 + 0.3749 = 0.8627 \end{aligned}$$

$$\begin{aligned} (\text{iii}) \quad P(X = 120) &= P(119.5 \leq X \leq 120.5) = P\left(\frac{19.5}{9.1} \leq z \leq \frac{20.5}{9.1}\right) \\ &= P(2.14 \leq z \leq 2.25) = P(0 \leq z \leq 2.25) - P(0 \leq z \leq 2.14) \\ &= 0.4878 - 0.4838 = 0.0040 \end{aligned}$$

Normal/ Approximation to Poisson Distribution

Normal distribution can also be used to approximate a Poisson distribution when its parameter $m \geq 10$. If X is a Poisson variate with mean m , then, for $m \geq 10$, the distribution of X can be taken as approximately normal with mean m and standard deviation \sqrt{m} so

that $z = \frac{X - m}{\sqrt{m}}$ is a standard normal variate.

Example 39: A random variable X follows Poisson distribution with parameter 25. Use normal approximation to Poisson distribution to find the probability that X is greater than or equal to 30.

Solution: $P(X \geq 30) = P(X \geq 29.5)$ (after making correction for continuity).

$$\begin{aligned} &= P\left(z \geq \frac{29.5 - 25}{5}\right) = P(z \geq 0.9) \\ &= 0.5000 - P(0 \leq z \leq 0.9) = 0.5000 - 0.3159 = 0.1841 \end{aligned}$$

Fitting a Normal Curve

A normal curve is fitted to the observed data with the following objectives :

1. To provide a visual device to judge whether it is a good fit or not.
2. Use to estimate the characteristics of the population.

The fitting of a normal curve can be done by

- (a) The Method of Ordinates or
- (b) The Method of Areas.

(a) **Method of Ordinates:** In this method, the ordinate $f(X)$ of the normal curve, for various values of the random variate X are obtained by using the table of ordinates for a standard normal variate.

$$\text{We can write } f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} = \frac{1}{\sigma} \phi(z)$$

$$\text{where } z = \frac{X - \mu}{\sigma} \text{ and } \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

The expected frequency corresponding to a particular value of X is given by

$y = N \cdot \phi(X) = \frac{N}{\sigma} \phi(z)$ and therefore, the expected frequency of a class $= y \times h$, where h is the class interval.

Example 40: Fit a normal curve to the following data :

Class Intervals : 10 - 20 20 - 30 30 - 40 40 - 50 50 - 60 60 - 70 70 - 80 Total								
Frequency	2	11	24	33	20	8	2	100

Solution: First we compute mean and standard deviation of the given data.

Class Intervals	Mid-values (X)	Frequency (f)	$d = \frac{X - 45}{10}$	fd	fd^2
10-20	15	2	-3	-6	18
20-30	25	11	-2	-22	44
30-40	35	24	-1	-24	24
40-50	45	33	0	0	0
50-60	55	20	1	20	20
60-70	65	8	2	16	32
70-80	75	2	3	6	18
Total		100		-10	156

Note: If the class intervals are not continuous, they should first be made so.

$$\therefore \mu = 45 - 10 \times \frac{10}{100} = 44$$

and $\sigma = 10 \sqrt{\frac{156}{100} - \left(\frac{10}{100}\right)^2} = 10\sqrt{1.55} = 12.4$

Table for the fitting of Normal Curve

Class Intervals	Mid-values (X)	$z = \frac{X - \mu}{\sigma}$	$\phi(z)$ (from table)	$y = \frac{N}{\sigma} \phi(z)$	$f_e *$
10-20	15	-2.34	0.0258	0.2081	2
20-30	25	-1.53	0.1238	0.9984	10
30-40	35	-0.73	0.3056	2.4645	25
40-50	45	0.08	0.3977	3.2073	32
50-60	55	0.89	0.2685	2.1653	22
60-70	65	1.69	0.0957	0.7718	8
70-80	75	2.50	0.0175	0.1411	1

(b) **Method of Areas:** Under this method, the probabilities or the areas of the random variable lying in various intervals are determined. These probabilities are then multiplied by N to get the expected frequencies. This procedure is explained below for the data of the above example.

Class Intervals	Lower Limit (X)	$z = \frac{X - 44}{12.4}$	Area from 0 to z	Area under the class	$f_e *$
10-20	10	-2.74	0.4969	0.0231	2
20-30	20	-1.94	0.4738	0.1030	10
30-40	30	-1.13	0.3708	0.2453	25
40-50	40	-0.32	0.1255	0.3099	31
50-60	50	0.48	0.1844	0.2171	22
60-70	60	1.29	0.4015	0.0806	8
70-80	70	2.10	0.4821	0.0160	2
	80	2.90	0.4981		

*Expected frequency approximated to the nearest integer.

Exercise with Hints

1. In a metropolitan city, there are on the average 10 fatal road accidents in a month (30 days). What is the probability that (i) there will be no fatal accident tomorrow, (ii) next fatal accident will occur within a week?

Hint: Take $m = 1/3$ and apply exponential distribution.

2. A counter at a super bazaar can entertain on the average 20 customers per hour. What is the probability that the time taken to serve a particular customer will be (i) less than 5 minutes, (ii) greater than 8 minutes?

Hint: Use exponential distribution.

3. The marks obtained in a certain examination follow normal distribution with mean 45 and standard deviation 10. If 1,000 students appeared at the examination, calculate the number of students scoring (i) less than 40 marks, (ii) more than 60 marks and (iii) between 40 and 50 marks.

Hint: See example 30.

4. The ages of workers in a large plant, with a mean of 50 years and standard deviation of 5 years, are assumed to be normally distributed. If 20% of the workers are below a certain age, find that age.

Hint: Given $P(X < X_1) = 0.20$, find X_1 .

5. The mean and standard deviation of certain measurements computed from a large sample are 10 and 3 respectively. Use normal distribution approximation to answer the following:

- (i) About what percentage of the measurements lie between 7 and 13 inclusive?
- (ii) About what percentage of the measurements are greater than 16?

Hint: Apply correction for continuity.

6. There are 600 business students in the post graduate department of a university and the probability for any student to need a copy of a particular text book from the university library on any day is 0.05. How many copies of the book should be kept in the library so that the probability that none of the students, needing a copy, has to come back disappointed is greater than 0.90? (Use normal approximation to binomial.)

Hint: If X_1 is the required number of copies, $P(X \leq X_1) \geq 0.90$.

7. The grades on a short quiz in biology were 0, 1, 2, 3, 10 points, depending upon the number of correct answers out of 10 questions. The mean grade was 6.7 with standard deviation of 1.2. Assuming the grades to be normally distributed, determine (i) the percentage of students scoring 6 points, (ii) the maximum grade of the lowest 10% of the class.

Hint: Apply normal approximation to binomial.

8. The following rules are followed in a certain examination. "A candidate is awarded a first division if his aggregate marks are 60% or above, a second division if his aggregate marks are 45% or above but less than 60% and a third division if the aggregate marks are 30% or above but less than 45%. A candidate is declared failed if his aggregate marks are below 30% and awarded a distinction if his aggregate marks are 80% or above."

At such an examination, it is found that 10% of the candidates have failed, 5% have obtained distinction. Calculate the percentage of students who were placed in the second division. Assume that the distribution of marks is normal. The areas under the standard normal curve from 0 to z are :

z	:	1.28	1.64	0.41	0.47
Area	:	0.4000	0.4500	0.1591	0.1808

Hint: First find parameters of the distribution on the basis of the given information.

9. For a certain normal distribution, the first moment about 10 is 40 and the fourth moment about 50 is 48. What is the mean and standard deviation of the distribution?

Hint: Use the condition $\beta_2 = 3$, for a normal distribution.

10. In a test of clerical ability, a recruiting agency found that the mean and standard deviation of scores for a group of fresh candidates were 55 and 10 respectively. For another experienced group, the mean and standard deviation of scores were found to be 62 and 8 respectively. Assuming a cut-off scores of 70, (i) what percentage of the experienced group is likely to be rejected, (ii) what percentage of the fresh group is likely to be selected, (iii) what will be the likely percentage of fresh candidates in the selected group? Assume that the scores are normally distributed.

Hint: See example 33.

11. 1,000 light bulbs with mean life of 120 days are installed in a new factory. Their length of life is normally distributed with standard deviation of 20 days. (i) How many bulbs will expire in less than 90 days? (ii) If it is decided to replace all the bulbs together, what interval should be allowed between replacements if not more than 10 percent bulbs should expire before replacement?

Hint: (ii) $P(X \leq X_1) = 0.9$.

12. The probability density function of a random variable is expressed as

$$p(X) = \sqrt{\left(\frac{2}{\pi}\right)} e^{-2(x-3)^2}, \quad (-\infty < X < \infty)$$

- (i) Identify the distribution.
- (ii) Determine the mean and standard deviation of the distribution.
- (iii) Write down two important properties of the distribution.

Hint: Normal distribution.

13. The weekly wages of 2,000 workers in a factory are normally distributed with a mean of Rs 200 and a variance of 400. Estimate the lowest weekly wages of the 197 highest paid workers and the highest weekly wages of the 197 lowest paid workers.

Hint: See example 32.

14. Among 10,000 random digits, in how many cases do we expect that the digit 3 appears at the most 950 times. (The area under the standard normal curve for $z = 1.667$ is 0.4525 approximately.)

Hint: $\mu = 10000 \times 0.10$ and $\sigma^2 = 1000 \times 0.9$.

15. Marks obtained by certain number of students are assumed to be normally distributed with mean 65 and variance 25. If three students are taken at random, what is the probability that exactly two of them will have marks over 70?

Hint: Find the probability (p) that a student gets more than 70 marks. Then find ${}^3C_2 p^2 q$.

16. The wage distribution of workers in a factory is normal with mean Rs 400 and standard deviation Rs 50. If the wages of 40 workers be less than Rs 350, what is the total number of workers in the factory? [given $\int_0^1 \phi(t) dt = 0.34$, where $t \sim N(0,1)$.]

Hint: $N \times \text{Probability that wage is less than } 350 = 40$.

17. The probability density function of a continuous random variable X is given by

$$f(X) = kX(2 - X), \quad 0 < X < 2 \\ = 0 \quad \text{elsewhere.}$$

Calculate the value of the constant k and E(X).

Hint: To find k, use the fact that total probability is unity.

18. $f(X) = \frac{5}{\sqrt{\pi}} e^{-25X^2}, \quad -\infty < X < \infty$ is the probability density function of a normal distribution with mean zero and variance 1/50. Is the statement true?

Hint: Transform X into standard normal variate z.

19. The income of a group of 10,000 persons was found to be normally distributed with mean Rs 1,750 p.m. and standard deviation Rs 50. Show that about 95% of the persons of the group had income exceeding Rs 1,668 and only 5% had income exceeding Rs 1,832.

Hint: See example 30.

20. A complex television component has 1,000 joints by a machine which is known to produce on an average one defective in forty. The components are examined and the faulty soldering corrected by hand. If the components requiring more than 35 corrections are discarded, what proportion of the components will be thrown away?

Hint: Use Poisson approximation to normal distribution.

21. The average number of units produced by a manufacturing concern per day is 355 with a standard deviation of 50. It makes a profit of Rs 150 per unit. Determine the percentage of days when its total profit per day is (i) between Rs 457.50 and Rs 645.00, (ii) greater than Rs 628.50.

Hint: Find the probabilities of producing 457.50/150 to 645/150 units etc.

22. A tyre manufacturing company wants 90% of its tyres to have a wear life of at least 40,000 kms. If the standard deviation of the wear lives is known to be 3,000 kms, find the lowest acceptable average wear life that must be achieved by the company. Assume that the wear life of tyres is normally distributed.

Hint: $P\left(\frac{40000 - \mu}{3000} > z_0\right) = 0.90.$

23. The average mileage before the scooter of a certain company needs a major overhaul is 60,000 kms with a S.D. of 10,000 kms. The manufacturer wishes to warranty these scooters, offering to make necessary overhaul free of charge if the buyer of a new scooter has a breakdown before covering certain number of kms. Assuming that the mileage, before an overhaul is required, is distributed normally, for how many kms should the manufacturer warranty so that not more than 3% of the new scooters come for free overhaul?

Hint: $P\left(\frac{X - 60000}{10000} < z_0\right) = 0.03.$

24. After an aeroplane has discharged its passengers, it takes crew A an average of 15 minutes ($\sigma = 4$ min.) to complete its task of handling baggage and loading food and other supplies. Crew B fuels the plane and does maintenance checks, taking an average of 16 minutes ($\sigma = 2$ min.) to complete its task. Assume that the two crews work independently and their times, to complete the tasks, are normally distributed. What is the probability that both crews will complete their tasks soon enough for the plane to be ready for take off with in 20 minutes?

Hint: $P(A).P(B) = P(AB).$

25. An automobile company buys nuts of a specified mean diameters m . A nut is classified as defective if its diameter differs from μ by more than 0.2 mm. The company requires that not more than 1% of the nuts may be defective. What should be the maximum variability that the manufacturer can allow in the production of nuts so as to satisfy the automobile company?

Hint: Find S.D.

Check Your Progress 11.2

1. What are the characteristics of Poisson Distribution?
2. What are the objectives for fitting a normal curve?

Notes: (a) Write your answer in the space given below.

- (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
- (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-
-

11.12 LET US SUM UP

Distribution	p.m.f./ p.d.f.	Range of R.V.	Parameters
(i) Binomial	${}^n C_r p^r q^{n-r}$	0,1,2,...,n	n and p
(ii) Hyper - geometric	$\frac{({}^k C_r)({}^{N-k} C_{n-r})}{{}^N C_r}$	0,1,2,...,n	n, N and k
(iii) Poisson	$\frac{e^{-m} \cdot m^r}{r!}$	0,1,2,..., ∞	m
(iv) Exponential	$m \cdot e^{-mt}$	$0 < t < \infty$	m
(v) Normal	$\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$	$-\infty < X < \infty$	μ and σ
(vi) S.Normal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2}$	$-\infty < z < \infty$	0 and 1

11.13 LESSON-END ACTIVITY

Apply the theoretical probability distribution to check the efficiency of industrial water treatment.

11.14 KEYWORDS

Binomial Distribution

Random Variable

Poisson Distribution

Normal Distribution

Exponential Distribution

Functions

11.15 QUESTIONS FOR DISCUSSION

1. Fill in the blanks:

- (a) Poisson distribution is used where no. of trials are
- (b) Poisson distribution is a distribution.
- (c) Normal distribution was first identified as
- (d) Normal probability curve is shaped.

2. Write True or False against each statement:

- (a) Binomial distribution assumption has Bernoulli Trials.
- (b) Standard deviation is equal to \sqrt{npq}
- (c) Poisson distribution is not used as a model.
- (d) The curve used to describe the accidental errors is Gaussian curve.

3. Write short notes on:

- (a) Fitting of Binomial Distribution
- (b) Pascal Distribution
- (c) Poisson Distribution
- (d) Geometrical Distribution
- (e) Normal Distribution

11.16 TERMINAL QUESTIONS

1. What do you understand by a theoretical probability distribution? How it is useful in business decision-making?
2. Define a binomial distribution. State the conditions under which binomial probability model is appropriate.
3. What are the parameters of a binomial distribution? Obtain expressions for mean and variance of the binomial variate in terms of these parameters.
4. What is a 'Poisson Process'? Obtain probability mass function of Poisson variate as a limiting form of the probability mass function of binomial variate.
5. Obtain mean and standard deviation of a Poisson random variate. Discuss some business and economic situations where Poisson probability model is appropriate.
6. How will you use Poisson distribution as an approximation to binomial? Explain with the help of an example.
7. Under what conditions will a random variable follow a normal distribution? State some important features of a normal probability curve.
8. What is a standard normal distribution? Discuss the importance of normal distribution in statistical theory.
9. State clearly the assumptions under which a binomial distribution tends to Poisson and to normal distribution.

10. Assume that the probability that a bomb dropped from an aeroplane will strike a target is $1/5$. If six bombs are dropped, find the probability that (i) exactly two will strike the target, (ii) at least two will strike the target.
11. An unbiased coin is tossed 5 times. Find the probability of getting (i) two heads, (ii) at least two heads.
12. An experiment succeeds twice as many times as it fails. Find the probability that in 6 trials there will be (i) no successes, (ii) at least 5 successes, (iii) at the most 5 successes.
13. In an army battalion 60% of the soldiers are known to be married and remaining unmarried. If $p(r)$ denotes the probability of getting r married soldiers from 5 soldiers, calculate $p(0), p(1), p(2), p(3), p(4)$ and $p(5)$. If there are 500 rows each consisting of 5 soldiers, approximately how many rows are expected to contain (i) all married soldiers, (ii) all unmarried soldiers?
14. A company has appointed 10 new secretaries out of which 7 are trained. If a particular executive is to get three secretaries, selected at random, what is the chance that at least one of them will be untrained?
15. The overall pass rate in a university examination is 70%. Four candidates take up such examination. What is the probability that (i) at least one of them will pass (ii) at the most 3 will pass (iii) all of them will pass, the examination?
16. 20% of bolts produced by a machine are defective. Deduce the probability distribution of the number of defectives in a sample of 5 bolts.
17. 25% employees of a firm are females. If 8 employees are chosen at random, find the probability that (i) 5 of them are males (ii) more than 4 are males (iii) less than 3 are females.
18. A supposed coffee connoisseur claims that he can distinguish between a cup of instant coffee and a cup of percolator coffee 75% of the times. It is agreed that his claim will be accepted if he correctly identifies at least 5 of the 6 cups. Find, (i) his chance of having the claim accepted if he is in fact only guessing and, (ii) his chance of having the claim rejected when he does not have the ability he claims.
19. It is known that 10% of the accounts of a firm contain errors. An auditor selects 5 accounts of the firm at random and finds that 3 of them contain errors. What is the probability of this result? What do you conclude on the basis of this result?
20. The incidence of an occupational disease in an industry is such that the workers have a 20% chance of suffering from it. What is the probability that out of 6 workmen, 4 or more will contract the disease?
21. A local politician claims that the assessed value of houses, for house tax purposes by the Municipal Corporation of Delhi, is not correct in 90% of the cases. Assuming this claim to be true, what is the probability that out of a sample of 4 houses selected at random (i) at least one will be found to be correctly assessed (ii) at least one will be found to be wrongly assessed?
22. There are 64 beds in a garden and 3 seeds of a particular type are sown in each bed. The probability of a flower being white is 0.25. Find the number of beds with 3, 2, 1 and 0 white flowers.
23. Suppose that half the population of a town are consumers of rice. 100 investigators are appointed to find out its truth. Each investigator interviews 10 individuals. How many investigators do you expect to report that three or less of the people interviewed are consumers of rice?
24. If the probability of a success in a trial is 0.2, find (a) mean, (b) variance, (c) moment coefficient of skewness and kurtosis of the number of successes in 100 trials.

25. There are 5 machines in a factory which may require adjustment from time to time during a day of their use. Two of these machines are of type I, each having a probability of $1/5$ of needing adjustment during a day and 3 are of type II, having corresponding probability of $1/10$.

Assuming that no machine needs adjustment twice on the same day, find the probability that on a particular day

- (i) Just 2 machines of type II and none of type I need adjustment.
- (ii) If just 2 machines need an adjustment, they are of the same type.

26. Fit a binomial distribution to the following data :

x	:	0	1	2	3	4
f	:	28	62	46	10	4

27. Five coins were tossed 100 times and the outcomes are recorded as given below. Compute the expected frequencies.

No. of heads	:	0	1	2	3	4	5
Observed frequency	:	2	10	24	38	18	8

28. The administrator of a large airport is interested in the number of aircraft departure delays that are attributable to inadequate control facilities. A random sample of 10 aircraft take off is to be thoroughly investigated. If the true proportion of such delays in all departures is 0.40, what is the probability that 4 of the sample departures are delayed because of control inadequacies? Also find mean, variance and mode of the random variable.
29. A company makes T.V.s., of which 15% are defective. 15 T.V.s. are shipped to a dealer. If each T.V. assembled is considered an independent trial, what is the probability that the shipment of 15 T.V.s. contain (i) no defective (ii) at the most one defective T.V.?
30. If 2% bulbs manufactured by a company are defective and the random variable denotes the number of defective bulbs, find mean, variance, measures of moment coefficient of skewness and kurtosis in a total of 50 bulbs.
31. 4096 families having just 4 children were chosen at random. Assuming the probability of a male birth equal to $1/2$, compute the expected number of families having 0, 1, 2, 3 and 4 male children.
32. If the number of telephone calls an operator receives from 9.00 A.M. to 9.30 A.M. follows a Poisson distribution with mean, $m = 2$, what is the probability that the operator will not receive a phone call in the same interval tomorrow?
33. (a) Write down the probability mass function of a Poisson distribution with parameter 3. What are the values of mean and variance of the corresponding random variable?
 (b) If X is a Poisson variate with parameter 2, find $P(3 \leq X \leq 5)$.
 (c) The standard deviation of a Poisson variate is 2, find $P(1 \leq X < 2)$.
34. Suppose that a telephone switch board handles 240 calls on the average during a rush hour and that the board can make at the most 10 connections per minute. Using Poisson distribution, estimate the probability that the board will be over taxed during a given minute.
35. An automatic machine makes paper clips from coils of a wire. On the average, 1 in 400 paper clips is defective. If the paper clips are packed in boxes of 100, what is the probability that any given box of clips will contain (i) no defective (ii) one or more defectives, (iii) less than two defectives?

36. Certain mass produced articles, of which 0.5% are defective, are packed in cartons each containing 100 articles. What proportions of the cartons are expected to be free from defective articles and what proportion contain, 2 or more defective articles?
37. A certain firm uses large fleet of delivery vehicles. Their records over a long period of time (during which their fleet size utilisation may be assumed to have remained constant) show that the average number of vehicles serviced per day is 3. Estimate the probability that on a given day
 - (i) no vehicle will be serviceable.
 - (ii) at the most 3 vehicles will be serviceable.
 - (iii) more than 2 vehicles will be unserviceable.
38. Suppose that a local electrical appliances shop has found from experience that the demand for tube lights is distributed as Poisson variate with a mean of 4 tube lights per week. If the shop keeps 6 tube lights during a particular week, what is the probability that the demand will exceed the supply during that week?
39. In a Poisson distribution, the probability of zero success is 15%. Find its mean and standard deviation.
40. Four unbiased coins are tossed 1,600 times. Using Poisson distribution, find the approximate probability of getting 4 heads r times.
41. The number of road accidents on a highway during a month follows a Poisson distribution with mean 6. Find the probability that in a certain month the number of accidents will be (i) not more than 3, (ii) between 2 and 4.
42. A random variable X follows Poisson law such that $P(X = k) = P(X = k + 1)$. Find its mean and variance.
43. The probability that a man aged 45 years will die within a year is 0.012. What is the probability that of 10 such men at least 9 will reach their 46th birthday? (Given $e^{-0.12} = 0.88692$).
44. During a period, persons arrive at a railway booking counter at the rate of 30 per hour. What is the probability that two or less persons will arrive in a period of 5 minutes?
45. An insurance company insures 4,000 people against loss of both eyes in a car accident. Based upon previous data, the rates were computed on the assumption that on the average 10 persons in 1,00,000 will have car accidents each year with this type of injury. What is the probability that more than 3 of the insured will collect on their policy in a given year?
46. A manufacturer, who produces medicine bottles, finds that 0.1% of the bottles are defective. The bottles are packed in boxes containing 500 bottles. A drug manufacturer buys 100 boxes from the producer of the bottles. Use Poisson distribution to find the number of boxes containing (i) no defective bottles (ii) at least two defective bottles.
47. A factory turning out lenses, supplies them in packets of 1,000. The packet is considered by the purchaser to be unacceptable if it contains 50 or more defective lenses. If a purchaser selects 30 lenses at random from a packet and adopts the criterion of rejecting the packet if it contains 3 or more defectives, what is the probability that the packet (i) will be accepted, (ii) will not be accepted?
48. 800 employees of a company are covered under the medical group insurance scheme. Under the terms of coverage, 40 employees are identified as belonging to 'High Risk' category. If 50 employees are selected at random, what is the probability that

(i) none of them is in the high risk category, (ii) at the most two are in the high risk category? (You may use Poisson approximation to Binomial).

49. In Delhi with 100 municipal wards, each having approximately the same population, the distribution of meningitis cases in 1987 were recorded as follows:

<i>No. of Cases</i>	: 0	1	2	3	4
<i>No. of Wards</i>	: 63	28	6	2	1

Fit a Poisson distribution to the above data.

50. The following table gives the number of days in 50 day-period during which automobile accidents occurred in a certain part of the city. Fit a Poisson distribution to the data.

<i>No. of accidents</i>	: 0	1	2	3	4
<i>No. of days</i>	: 19	18	8	4	1

51. A sample of woollen balls has a mean weight of 3.2 oz. and standard deviation of 1 oz. Assuming that the weight of woollen balls is distributed normally, (i) How many balls are expected to weigh between 2.7 and 3.7 oz., (ii) what is the probability that weight of a ball is less than 1.5 oz. and (iii) what is the probability that the weight of the ball will exceed 4.7 oz.?
52. The weekly wages of 2,000 workers are normally distributed. Its mean and standard deviation are Rs 140 and Rs 10 respectively. Estimate the number of workers with weekly wages
- (i) Between Rs 120 and Rs 130.
 - (ii) More than Rs 170.
 - (iii) Less than Rs 165.
53. Find the probability that the value of an item drawn at random from a normal distribution with mean 20 and standard deviation 10 will be between (i) 10 and 15, (ii) – 5 and 10 and (iii) 15 and 25.
54. In a manufacturing organisation the distribution of wages was perfectly normal and the number of workers employed in the organisation were 5,000. The mean wages of the workers were calculated as Rs 800 p.m. and the standard deviation was worked out to be Rs 200. Estimate
- (i) the number of workers getting wages between Rs 700 and Rs 900.
 - (ii) the percentage of workers getting wages above Rs 1,000.
 - (iii) the percentage of workers getting wages below Rs 600.
55. Suppose that the waist measurements W of 800 girls are normally distributed with mean 66 cms and standard deviation 5 cms. Find the number of girls with waists;
- (i) between 65 and 70 cms. (ii) greater than or equal to 72 cms.
56. (a) A normal distribution has 77.0 as its mean. Find its standard deviation if 20% of the area under the curve lies to the right of 90.0.
- (b) A random variable has a normal distribution with 10 as its standard deviation. Find its mean if the probability that the random variable takes on a value less than 80.5 is 0.3264.
57. In a particular examination an examinee can get marks ranging from 0 to 100. Last year, 1,00,000 students took this examination. The marks obtained by them followed a normal distribution. What is the probability that the marks obtained by a student selected at random would be exactly 63?

58. A collection of human skulls is divided into three classes according to the value of a 'length breadth index' x . Skulls with $x < 75$ are classified as 'long', those with $75 < x < 80$ as 'medium' and those with $x > 80$ as 'short'. The percentage of skulls in the three classes in this collection are respectively 58, 38 and 4. Find, approximately, the mean and standard deviation of x on the assumption that it is normally distributed.
59. A wholesale distributor of a fertiliser product finds that the annual demand for one type of fertiliser is normally distributed with a mean of 120 tonnes and standard deviation of 16 tonnes. If he orders only once a year, what quantity should be ordered to ensure that there is only a 5% chance of running short?
60. A multiple choice quiz has 200 questions, each with 4 possible answers of which only one is correct. What is the probability (using normal approximation to binomial distribution) that sheer guess work yields from 25 to 30 correct answers for 80 questions (out of 200 questions) about which the student has no knowledge?
61. In a normal distribution 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution.
62. The mean life of the bulbs manufactured by a company is estimated to be 2,025 hours. By using normal approximation to Poisson distribution, estimate the percentage of bulbs that are expected to last for (i) less than 2,100 hours, (ii) between 1,900 and 2,000 hours and (iii) more than 2,000 hours.
63. Find mean and standard deviation if a score of 51 is 2 standard deviation above mean and a score of 42 is 1 standard deviation below mean. Assume that the scores are normally distributed.
64. (a) A manufacturer requires washers with specification of its inner diameter as 3.30 ± 0.04 mm. If the inner diameters of the washers supplied by some suppliers are distributed normally with mean $\mu = 3.31$ mm. and $\sigma = 0.02$ mm., what percentage of the washers, supplied in the a lot, are expected to meet the required specification?
 (b) A cylinder making machine has $\sigma = 0.5$ mm. At what value of m should the machine be set to ensure that 2.5% of the cylinders have diameters of 25.48 mm. or more?
65. The mean life of a pair of shoes manufactured by a company is estimated to be 2.59 years with a standard deviation of 3 months. What should be fixed as guarantee period so that the company has not to replace more than 5% of the pairs?
66. In a large group of men, it is found that 5% are under 60 inches and 40% are between 60 and 65 inches in height. Assuming the distribution to be exactly normal, find the mean and standard deviation of the height. The values of z for area equal to 0.45 and 0.05 between 0 to z are 1.645 and 0.125 respectively.
67. Packets of a certain washing powder are filled with an automatic machine with an average weight of 5 kg. and a standard deviation of 50 gm. If the weights of packets are normally distributed, find the percentage of packets having weight above 5.10 kg.
68. For a normal distribution with mean 3 and variance 16, find the value of y such that the probability of the variate lying in the interval (3, y) is 0.4772.
69. The mean income of people working in an industrial city is approximated by a normal distribution with a mean of Rs 24,000 and a standard deviation of Rs 3,000. What percentage of the people in this city have income exceeding Rs 28,500? In a random sample of 50 employed persons of this city, about how many can be expected to have income less than Rs 19,500?

70. The burning time of an experimental rocket is a random variable which has normal distribution with $\mu = 4.36$ seconds and $s = 0.04$ seconds. What are the probabilities that this kind of rocket will burn for
 - (i) less than 4.5 seconds, (ii) more than 4.40 seconds, (iii) between 4.30 to 4.42 seconds.
71. A company manufactures batteries and guarantees them for a life of 24 months.
 - (i) If the average life has been found in tests to be 33 months with a standard deviation of 4 months, how many batteries will have to be replaced under guarantee if the life of the batteries follows a normal distribution?
 - (ii) If annual sales are 10,000 batteries at a profit of Rs 50 each and each replacement costs the company Rs 100, find the net profit.
 - (iii) Would it be worth its while to extend the guarantee to 27 months if the sales were to be increased by this extra offer to 12,000 batteries?
72. The distribution of total time a light bulb will last from the moment it is first put into service is known to be exponential with mean time between failure of the bulbs equal to 1,000 hours. What is the probability that the bulb will last for more than 1,000 hours?
73. An editor of a publishing company calculates that it requires 11 months on an average to complete the publication process from the manuscript to finished book with a standard deviation of 2.4 months. He believes that the distribution of publication time is well described by a normal distribution. Determine, out of 190 books that he will handle this year, how many will complete the process in less than a year?
74. The I.Q.'s of army volunteers in a given year are normally distributed with mean = 110 and standard deviation = 10. The army wants to give advanced training to 20% of those recruits with the highest scores. What is the lowest I.Q. score acceptable for the advanced training?
75. If 60% of the voters in a constituency favour a particular candidate, find the probability that in a sample of 300 voters, more than 170 voters would favour the candidate. Use normal approximation to the binomial.
76. From the past experience, a committee for admission to certain course consisting of 200 seats, has estimated that 5% of those granted admission do not turn up. If 208 letters of intimation of admission are issued, what is the probability that seat is available for all those who turn up? Use normal approximation to the binomial.
77. The number of customer arrivals at a bank is a Poisson process with average of 6 customers per 10 minutes. (a) What is the probability that the next customer will arrive within 3 minutes? (b) What is the probability that the time until the next customer arrives will be from 2 to 3 minutes? (c) What is the probability that the next customer will arrive after more than 4 minutes?
78. Comment on the following statements :
 - (i) The mean of a normal distribution is 10 and the third order central moment is 1.5.
 - (ii) The mean of a Poisson variate is 4 and standard deviation is $\sqrt{3}$.
 - (iii) The mean of a binomial variate is 10 and standard deviation is 4.
 - (iv) The probability that a discrete random variable X takes a value $X = a$ is equal to $P(X = a)$, where $P(X)$ is probability mass function of the random variable.
 - (v) The probability that a continuous random variable X takes a value $X = a$ is equal to $f(X = a)$, where $f(X)$ is probability density function of the random variable.

- (vi) The second raw moment of a Poisson distribution is 2. The probability $P(X = 0) = e^{-1}$.
- (vii) The variance of a binomial distribution cannot exceed $\frac{n}{4}$.
- (viii) If for a Poisson variate X , $P(X = 1) = P(X = 2)$, then $E(X) = 2$.
- (ix) If for a Poisson variate X , $P(X = 0) = P(X = 1)$, then $P(X > 0) = e^{-1}$.
- (x) $\beta_1 = 0$ and $\beta_2 = 3$ for a normal distribution.
79. State whether the following statements are True/False :
- Mean of a binomial variate is always greater than its variance.
 - Mean of a Poisson variate may or may not be equal to its variance.
 - Time required for the arrival of two telephone calls at a desk is a Poisson variate.
 - A normal distribution is always symmetrical.
 - A binomial distribution with $p = q$ is always symmetrical.
 - The probability function of a continuous random variable is called a probability mass function.
 - The parameters of a distribution completely determine the distribution.
 - Any normal variate with given mean and standard deviation can be transformed into a standard normal variate.
 - The number of suicide cases in a given year is a binomial variate.
 - Since the probability that a continuous random variable takes a particular value is zero, the event is said to be impossible.
80. ***Fill in the blanks :***
- If three balls are drawn, successively with replacement, from a bag containing 4 red and 3 black balls, the number of red balls is a random variable.
 - A standard normal variate has mean equal to and standard deviation equal to
 - When $1 - p > p$, the binomial distribution is skewed.
 - The of a binomial variate with mean = 4 and standard deviation = $\sqrt{3}$ are 16 and $\frac{1}{4}$.
 - A normal variate obtained by subtracting its mean and dividing by its standard deviation is known as variate.
 - If the expected value of a Poisson variate is 9, its is 3.
 - The number of defects per unit of length of a wire is a variate.
 - The time of occurrence of an event is an variate.
 - The number of trials needed to get a given number of successes is a variate.
 - Normal distribution is also known as the normal law of

11.17 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

1. (a) large (b) discrete (c) law of errors (d) Bell
2. (a) True (b) True (c) False (d) True

11.18 SUGGESTED READINGS

Peyton Z. Peebles, Jr., *Probability Random Variables and Random Signal Principles*.

Harold Crames, *Random variables and probability distributions*, Cambridge Univ. Press.

M.A. Kon, *Probability Distributions in Quantum Statistical Mechanics*. Springer Verlag (Jan. 1966)

Charles F. Manski, *Probability Identification of Probability Distribution*, Springer Verlag (May 2003)

P.C. Consul, Femix Farieye Legrargian, *Probability Distributions*, Birkhauses, Jan. 2006.

LESSON

12

PROBABILITY DISTRIBUTION OF A RANDOM VARIABLE

CONTENTS

- 12.0 Aims and Objectives
 - 12.1 Introduction
 - 12.2 Probability Distribution of a Random Variable
 - 12.2.1 Discrete and Continuous Probability Distributions
 - 12.2.2 Cumulative Probability Function or Distribution Function
 - 12.3 Mean and Variance of a Random Variable
 - 12.4 Theorems on Expectation
 - 12.5 Joint Probability Distribution
 - 12.5.1 Marginal Probability Distribution
 - 12.5.2 Conditional Probability Distribution
 - 12.5.3 Expectation of the Sum or Product of two Random Variables
 - 12.5.4 Expectation of a Function of Random Variables
 - 12.6 Decision Analysis under Certainty
 - 12.7 Decision-making under Uncertainty
 - 12.8 Decision-making under Risk
 - 12.9 Expected Value with Perfect Information (EVPI)
 - 12.10 Use of Subjective Probabilities in Decision-making
 - 12.11 Use of Posterior Probabilities in Decision-making
 - 12.12 Let us Sum Up
 - 12.13 Lesson-end Activity
 - 12.14 Keywords
 - 12.15 Questions for Discussion
 - 12.16 Terminal Questions
 - 12.17 Model Answers to Questions for Discussion
 - 12.18 Suggested Readings
-

12.0 AIMS AND OBJECTIVES

In the previous two lessons we had talked about the probability theory and various distributions. But the most important aspect is to have a flavour of random variable so that it may be attributed to a function defined on a state space.

12.1 INTRODUCTION

A random variable X is a real valued function of the elements of sample space S , i.e., different values of the random variable are obtained by associating a real number with each element of the sample space. A random variable is also known as a *stochastic* or *chance variable*.

Mathematically, we can write $X = F(e)$, where $e \in S$ and X is a real number. We can note here that the domain of this function is the set S and the range is a set or subset of real numbers.

Example 1: Three coins are tossed simultaneously. Write down the sample space of the random experiment. What are the possible values of the random variable X , if it denotes the number of heads obtained?

Solution: The sample space of the experiment can be written as

$$S = \{(H,H,H), (H,H,T), (H,T,H), (T,H,H), (H,T,T), (T,H,T), (T,T,H), (T,T,T)\}$$

We note that the first element of the sample space denotes 3 heads, therefore, the corresponding value of the random variable will be 3. Similarly, the value of the random variable corresponding to each of the second, third and fourth element will be 2 and it will be 1 for each of the fifth, sixth and seventh element and 0 for the last element. Thus, the random variable X , defined above can take four possible values, i.e., 0, 1, 2 and 3.

It may be pointed out here that it is possible to define another random variable on the above sample space.

12.2 PROBABILITY DISTRIBUTION OF A RANDOM VARIABLE

Given any random variable, corresponding to a sample space, it is possible to associate probabilities to each of its possible values. For example, in the toss of 3 coins, assuming that they are unbiased, the probabilities of various values of the random variable X , defined in example 1 above, can be written as :

$$P(X=0) = \frac{1}{8}, \quad P(X=1) = \frac{3}{8}, \quad P(X=2) = \frac{3}{8} \text{ and } P(X=3) = \frac{1}{8}.$$

The set of all possible values of the random variable X alongwith their respective probabilities is termed as Probability Distribution of X . The probability distribution of X , defined in example 1 above, can be written in a tabular form as given below:

X	:	0	1	2	3	Total
$p(X)$:	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Note that the total probability is equal to unity.

In general, the set of n possible values of a random variable X , i.e., $\{X_1, X_2, \dots, X_n\}$ along with their respective probabilities $p(X_1), p(X_2), \dots, p(X_n)$, where

$\sum_{i=1}^n p(X_i) = 1$, is called a probability distribution of X . The expression $p(X)$ is called the probability function of X .

12.2.1 Discrete and Continuous Probability Distributions

Like any other variable, a random variable X can be discrete or continuous. If X can take only finite or countably infinite set of values, it is termed as a discrete random variable. On the other hand, if X can take an uncountable set of infinite values, it is called a continuous random variable.

The random variable defined in example 1 is a discrete random variable. However, if X denotes the measurement of heights of persons or the time interval of arrival of a specified number of calls at a telephone desk, etc., it would be termed as a continuous random variable.

The distribution of a discrete random variable is called the Discrete Probability Distribution and the corresponding probability function $p(X)$ is called a Probability Mass Function. In order that any discrete function $p(X)$ may serve as probability function of a discrete random variable X , the following conditions must be satisfied :

$$(i) \quad p(X_i) \geq 0 \quad \forall i = 1, 2, \dots, n \text{ and}$$

$$(ii) \quad \sum_{i=1}^n p(X_i) = 1$$

In a similar way, the distribution of a continuous random variable is called a *Continuous Probability Distribution* and the corresponding probability function $p(X)$ is termed as the *Probability Density Function*. The conditions for any function of a continuous variable to serve as a probability density function are :

$$(i) \quad p(X) \geq 0 \quad \forall \text{ real values of } X, \text{ and}$$

$$(ii) \quad \int_{-\infty}^{\infty} p(X) dX = 1$$

Remarks:

1. When X is a continuous random variable, there are an infinite number of points in the sample space and thus, the probability that X takes a particular value is always defined to be zero even though the event is not regarded as impossible. Hence, we always talk of the probability of a continuous random variable lying in an interval.
2. The concept of a probability distribution is not new. In fact it is another way of representing a frequency distribution. Using statistical definition, we can treat the relative frequencies of various values of the random variable as the probabilities.

Example 2: Two unbiased die are thrown. Let the random variable X denote the sum of points obtained. Construct the probability distribution of X .

Solution: The possible values of the random variable are :

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$$

The probabilities of various values of X are shown in the following table :

Probability Distribution of X

X	2	3	4	5	6	7	8	9	10	11	12	Total
$p(X)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

Example 3: Three marbles are drawn at random from a bag containing 4 red and 2 white marbles. If the random variable X denotes the number of red marbles drawn, construct the probability distribution of X .

Solution: The given random variable can take 3 possible values, i.e., 1, 2 and 3. Thus, we can compute the probabilities of various values of the random variable as given below :

$$P(X = 1, \text{i.e., 1R and 2W marbles are drawn}) = \frac{{}^4C_1 \times {}^2C_2}{{}^6C_3} = \frac{4}{20}$$

$$P(X = 2, \text{ i.e., } 2R \text{ and } 1W \text{ marbles are drawn}) = \frac{{}^4C_2 \times {}^2C_1}{{}^6C_3} = \frac{12}{20}$$

$$P(X = 3, \text{ i.e., } 3R \text{ marbles are drawn}) = \frac{{}^4C_3}{{}^6C_3} = \frac{4}{20}$$

Note: In the event of white balls being greater than 2, the possible values of the random variable would have been 0, 1, 2 and 3.

12.2.2 Cumulative Probability Function or Distribution Function

This concept is similar to the concept of cumulative frequency. The distribution function is denoted by $F(x)$.

For a discrete random variable X , the distribution function or the cumulative probability function is given by $F(x) = P(X \leq x)$.

If X is a random variable that can take values, say 0, 1, 2,, then

$$F(1) = P(X = 0) + P(X = 1), F(2) = P(X = 0) + P(X = 1) + P(X = 2), \text{ etc.}$$

Similarly, if X is a continuous random variable, the distribution function or cumulative probability density function is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^x p(X)dX$$

12.3 MEAN AND VARIANCE OF A RANDOM VARIABLE

The mean and variance of a random variable can be computed in a manner similar to the computation of mean and variance of the variable of a frequency distribution.

Mean: If X is a discrete random variable which can take values X_1, X_2, \dots, X_n , with respective probabilities as $p(X_1), p(X_2), \dots, p(X_n)$, then its mean, also known as the *Mathematical Expectation* or *Expected Value* of X , is given by :

$$E(X) = X_1p(X_1) + X_2p(X_2) + \dots + X_np(X_n) = \sum_{i=1}^n X_i p(X_i).$$

The mean of a random variable or its probability distribution is often denoted by μ , i.e., $E(X) = \mu$.

Remarks: The mean of a frequency distribution can be written as

$$X_1 \cdot \frac{f_1}{N} + X_2 \cdot \frac{f_2}{N} + \dots + X_n \cdot \frac{f_n}{N}, \text{ which is identical to the expression for expected value.}$$

Variance: The concept of variance of a random variable or its probability distribution is also similar to the concept of the variance of a frequency distribution.

The variance of a frequency distribution is given by

$$\sigma^2 = \frac{1}{N} \sum f_i (X_i - \bar{X})^2 = \sum (X_i - \bar{X})^2 \cdot \frac{f_i}{N} = \text{Mean of } (X_i - \bar{X})^2 \text{ values.}$$

The expression for variance of a probability distribution with mean M can be written in a similar way, as given below :

$$\sigma^2 = E(X - \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2 p(X_i), \text{ where } X \text{ is a discrete random variable.}$$

Remarks: If X is a continuous random variable with probability density function p(X), then

Probability Distribution
of a Random Variable

$$E(X) = \int_{-\infty}^{\infty} X \cdot p(X) dX$$

$$\sigma^2 = E(X - \mu)^2 = \int_{-\infty}^{\infty} (X - \mu)^2 \cdot p(X) dX$$

Moments

The r th moment of a discrete random variable about its mean is defined as:

$$\mu_r = E(X - \mu)^r = \sum_{i=1}^n (X_i - \mu)^r p(X_i)$$

Similarly, the r th moment about any arbitrary value A, can be written as

$$\mu'_r = E(X - A)^r = \sum_{i=1}^n (X_i - A)^r p(X_i)$$

The expressions for the central and the raw moments, when X is a continuous random variable, can be written as

$$\mu_r = E(X - \mu)^r = \int_{-\infty}^{\infty} (X - \mu)^r \cdot p(X) dX$$

and $\mu'_r = E(X - A)^r = \int_{-\infty}^{\infty} (X - A)^r \cdot p(X) dX$ respectively.

12.4 THEOREMS ON EXPECTATION

Theorem 1: Expected value of a constant is the constant itself, i.e., $E(b) = b$, where b is a constant.

Proof: The given situation can be regarded as a probability distribution in which the random variable takes a value b with probability 1 and takes some other real value, say a, with probability 0.

Thus, we can write $E(b) = b \times 1 + a \times 0 = b$

Theorem 2: $E(aX) = aE(X)$, where X is a random variable and a is constant.

Proof: For a discrete random variable X with probability function $p(X)$, we have :

$$E(aX) = aX_1 \cdot p(X_1) + aX_2 \cdot p(X_2) + \dots + aX_n \cdot p(X_n)$$

$$= a \sum_{i=1}^n X_i \cdot p(X_i) = aE(X)$$

Combining the results of theorems 1 and 2, we can write

$$E(aX + b) = aE(X) + b$$

Remarks : Using the above result, we can write an alternative expression for the variance of X, as given below :

$$\begin{aligned} \sigma^2 &= E(X - \mu)^2 = E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 = E(X^2) - [E(X)]^2 \\ &= \text{Mean of Squares} - \text{Square of the Mean} \end{aligned}$$

We note that the above expression is identical to the expression for the variance of a frequency distribution.

Theorems on Variance

Theorem 1: The variance of a constant is zero.

Proof: Let b be the given constant. We can write the expression for the variance of b as:

$$\text{Var}(b) = E[b - E(b)]^2 = E[b - b]^2 = 0.$$

Theorem 2: $\text{Var}(X + b) = \text{Var}(X)$.

Proof: We can write $\text{Var}(X + b) = E[X + b - E(X + b)]^2 = E[X + b - E(X) - b]^2$
 $= E[X - E(X)]^2 = \text{Var}(X)$

Similarly, it can be shown that $\text{Var}(X - b) = \text{Var}(X)$

Remarks: The above theorem shows that variance is independent of change of origin.

Theorem 3: $\text{Var}(aX) = a^2\text{Var}(X)$

Proof: We can write $\text{Var}(aX) = E[aX - E(aX)]^2 = E[aX - aE(X)]^2$
 $= a^2E[X - E(X)]^2 = a^2\text{Var}(X).$

Combining the results of theorems 2 and 3, we can write

$$\text{Var}(aX + b) = a^2\text{Var}(X).$$

This result shows that the variance is independent of change of origin but not of change of scale.

Remarks:

1. On the basis of the theorems on expectation and variance, we can say that if X is a random variable, then its linear combination, $aX + b$, is also a random variable with mean $aE(X) + b$ and Variance equal to $a^2\text{Var}(X)$.
2. The above theorems can also be proved for a continuous random variable.

Example 4: Compute mean and variance of the probability distributions obtained in examples 1, 2 and 3.

Solution:

- (a) The probability distribution of X in example 1 was obtained as

X	0	1	2	3
$p(X)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

From the above distribution, we can write

$$E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1.5$$

To find variance of X , we write

$$\text{Var}(X) = E(X^2) - [E(X)]^2, \text{ where } E(X^2) = \sum X^2 p(X).$$

$$\text{Now, } E(X^2) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 4 \times \frac{3}{8} + 9 \times \frac{1}{8} = 3$$

$$\text{Thus, } \text{Var}(X) = 3 - (1.5)^2 = 0.75$$

- (b) The probability distribution of X in example 2 was obtained as

X	2	3	4	5	6	7	8	9	10	11	12	Total
$p(X)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

$$\therefore E(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36} \\ + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = \frac{252}{36} = 7$$

$$\text{Further, } E(X^2) = 4 \times \frac{1}{36} + 9 \times \frac{2}{36} + 16 \times \frac{3}{36} + 25 \times \frac{4}{36} + 36 \times \frac{5}{36} + 49 \times \frac{6}{36} \\ + 64 \times \frac{5}{36} + 81 \times \frac{4}{36} + 100 \times \frac{3}{36} + 121 \times \frac{2}{36} + 144 \times \frac{1}{36} = \frac{1974}{36} = 54.8$$

$$\text{Thus, } \text{Var}(X) = 54.8 - 49 = 5.8$$

(c) The probability distribution of X in example 3 was obtained as

X	1	2	3
$p(X)$	$\frac{4}{20}$	$\frac{12}{20}$	$\frac{4}{20}$

From the above, we can write

$$E(X) = 1 \times \frac{4}{20} + 2 \times \frac{12}{20} + 3 \times \frac{4}{20} = 2$$

$$\text{and } E(X^2) = 1 \times \frac{4}{20} + 4 \times \frac{12}{20} + 9 \times \frac{4}{20} = 4.4$$

$$\therefore \text{Var}(X) = 4.4 - 4 = 0.4$$

Expected Monetary Value (EMV)

When a random variable is expressed in monetary units, its expected value is often termed as expected monetary value and symbolised by EMV.

Example 5: If it rains, an umbrella salesman earns Rs 100 per day. If it is fair, he loses Rs 15 per day. What is his expectation if the probability of rain is 0.3?

Solution: Here the random variable X takes only two values, $X_1 = 100$ with probability 0.3 and $X_2 = -15$ with probability 0.7.

Thus, the expectation of the umbrella salesman

$$= 100 \times 0.3 - 15 \times 0.7 = 19.5$$

The above result implies that his average earning in the long run would be Rs 19.5 per day.

Example 6: A person plays a game of throwing an unbiased die under the condition that he could get as many rupees as the number of points obtained on the die. Find the expectation and variance of his winning. How much should he pay to play in order that it is a fair game?

Solution: The probability distribution of the number of rupees won by the person is given below :

$X(\text{Rs})$	1	2	3	4	5	6
$p(X)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\text{Thus, } E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \text{Rs } \frac{7}{2}$$

$$\text{and } E(X^2) = 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6} + 16 \times \frac{1}{6} + 25 \times \frac{1}{6} + 36 \times \frac{1}{6} = \frac{91}{6}$$

$$\therefore \sigma^2 = \frac{91}{6} - \left(\frac{7}{2} \right)^2 = \frac{35}{12} = 2.82. \text{ Note that the unit of } s^2 \text{ will be (Rs)}^2.$$

Since $E(X)$ is positive, the player would win Rs 3.5 per game in the long run. Such a game is said to be favourable to the player. In order that the game is fair, the expectation of the player should be zero. Thus, he should pay Rs 3.5 before the start of the game so that the possible values of the random variable become $1 - 3.5 = -2.5$, $2 - 3.5 = -1.5$, $3 - 3.5 = -0.5$, $4 - 3.5 = 0.5$, etc. and their expected value is zero.

Example 7: Two persons A and B throw, alternatively, a six faced die for a prize of Rs 55 which is to be won by the person who first throws 6. If A has the first throw, what are their respective expectations?

Solution: Let A be the event that A gets a 6 and B be the event that B gets a 6. Thus,

$$P(A) = \frac{1}{6} \text{ and } P(B) = \frac{1}{6}.$$

If A starts the game, the probability of his winning is given by :

$$\begin{aligned} P(A \text{ wins}) &= P(A) + P(\bar{A}).P(\bar{B}).P(A) + P(\bar{A}).P(\bar{B}).P(\bar{A}).P(\bar{B}).P(A) + \dots \\ &= \frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + \dots \\ &= \frac{1}{6} \left[1 + \left(\frac{5}{6} \right)^2 + \left(\frac{5}{6} \right)^4 + \dots \right] = \frac{1}{6} \times \left(\frac{1}{1 - \frac{25}{36}} \right) = \frac{1}{6} \times \frac{36}{11} = \frac{6}{11} \end{aligned}$$

Similarly, $P(B \text{ wins}) = P(\bar{A}).P(B) + P(\bar{A}).P(\bar{B}).P(\bar{A}).P(B) + \dots$

$$\begin{aligned} &= \frac{5}{6} \times \frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + \dots \\ &= \frac{5}{6} \times \frac{1}{6} \left[1 + \left(\frac{5}{6} \right)^2 + \left(\frac{5}{6} \right)^4 + \dots \right] = \frac{5}{6} \times \frac{1}{6} \times \frac{36}{11} = \frac{5}{11} \end{aligned}$$

Expectation of A and B for the prize of Rs 55

Since the probability that A wins is $\frac{6}{11}$, therefore, the random variable takes a value 55

with probability $\frac{6}{11}$ and value 0 with probability $\frac{5}{11}$. Hence, $E(A) = 55 \times \frac{6}{11} + 0 \times \frac{5}{11} = \text{Rs } 30$

Similarly, the expectation of B is given by $E(B) = 55 \times \frac{6}{11} + 0 \times \frac{5}{11} = \text{Rs } 30$

Example 8: An unbiased die is thrown until a four is obtained. Find the expected value and variance of the number of throws.

Solution: Let X denote the number of throws required to get a four. Thus, X will take values 1, 2, 3, 4, with respective probabilities.

$$\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6} \right)^2 \times \frac{1}{6}, \left(\frac{5}{6} \right)^3 \times \frac{1}{6} \dots \text{etc.}$$

$$\begin{aligned} \therefore E(X) &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{5}{6} \cdot \frac{1}{6} + 3 \cdot \left(\frac{5}{6} \right)^2 \cdot \frac{1}{6} + 4 \cdot \left(\frac{5}{6} \right)^3 \cdot \frac{1}{6} \dots \\ &= \frac{1}{6} \left[1 + 2 \cdot \frac{5}{6} + 3 \cdot \left(\frac{5}{6} \right)^2 + 4 \cdot \left(\frac{5}{6} \right)^3 + \dots \right] \end{aligned}$$

$$\text{Let } S = 1 + 2 \cdot \frac{5}{6} + 3 \cdot \left(\frac{5}{6} \right)^2 + 4 \cdot \left(\frac{5}{6} \right)^3 + \dots$$

Multiplying both sides by $\frac{5}{6}$, we get

$$\begin{aligned}\frac{5}{6}S &= \frac{5}{6} + 2\left(\frac{5}{6}\right)^2 + 3\left(\frac{5}{6}\right)^3 + 4\left(\frac{5}{6}\right)^4 + \dots \\ \therefore S - \frac{5}{6}S &= 1 + (2-1)\frac{5}{6} + (3-2)\left(\frac{5}{6}\right)^2 + (4-3)\left(\frac{5}{6}\right)^3 + \dots \\ \frac{1}{6}S &= 1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^3 + \dots = \frac{1}{1-\frac{5}{6}} = 6\end{aligned}\quad \dots (1)$$

Thus, $S = 36$ and hence $E(X) = \frac{1}{6} \times 36 = 6$.

Further, to find variance, we first find $E(X^2)$

$$\begin{aligned}E(X^2) &= 1 \cdot \frac{1}{6} + 2^2 \cdot \frac{5}{6} \cdot \frac{1}{6} + 3^2 \cdot \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6} + 4^2 \cdot \left(\frac{5}{6}\right)^3 \cdot \frac{1}{6} + \dots \\ &= \frac{1}{6} \left[1 + 2^2 \cdot \left(\frac{5}{6}\right)^2 + 3^2 \cdot \left(\frac{5}{6}\right)^3 + 4^2 \cdot \left(\frac{5}{6}\right)^4 + \dots \right]\end{aligned}$$

$$\text{Let } S = 1 + 2^2 \cdot \left(\frac{5}{6}\right)^2 + 3^2 \cdot \left(\frac{5}{6}\right)^3 + 4^2 \cdot \left(\frac{5}{6}\right)^4 + \dots$$

Multiply both sides by $\frac{5}{6}$ and subtract from S , to get

$$\begin{aligned}\frac{1}{6}S &= 1 + (2^2 - 1)\left(\frac{5}{6}\right) + (3^2 - 2^2)\left(\frac{5}{6}\right)^2 + (4^2 - 3^2)\left(\frac{5}{6}\right)^3 + \dots \\ &= 1 + 3\left(\frac{5}{6}\right) + 5\left(\frac{5}{6}\right)^2 + 7\left(\frac{5}{6}\right)^3 + \dots\end{aligned}$$

Further, multiply both sides by $\frac{5}{6}$ and subtract

$$\begin{aligned}\frac{1}{6}S - \frac{5}{36}S &= 1 + (3-1)\left(\frac{5}{6}\right) + (5-3)\left(\frac{5}{6}\right)^2 + (7-5)\left(\frac{5}{6}\right)^3 + \dots \\ \frac{1}{36}S &= 1 + 2\left(\frac{5}{6}\right) \left\{ 1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2 + \dots \right\} = 1 + \frac{5}{3} \times 6 = 11\end{aligned}\quad \dots (2)$$

$$\therefore S = 36 \times 11 \text{ and } E(X^2) = \frac{1}{6} \times 36 \times 11 = 66$$

Hence, Variance = $E(X^2) - [E(X)]^2 = 66 - 36 = 30$

Generalisation: Let p be the probability of getting 4, then from equation (1) we can

$$\text{write } pS = \frac{1}{1-q} = \frac{1}{p} \text{ or } S = \frac{1}{p^2} \text{ Therefore, } E(X) = p \left(\frac{1}{p^2} \right) = \frac{1}{p}$$

Similarly, equation (2) can be written as

$$p^2S = 1 + \frac{2q}{p} \text{ or } S = \frac{1}{p^2} + \frac{2q}{p^3} = \frac{p+2q}{p^3}$$

$$\text{Therefore, } E(X^2) = p \left(\frac{p+2q}{p^3} \right) = \frac{p+2q}{p^2} \text{ and } \text{Var}(X) = \frac{p+2q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$$

12.5 JOINT PROBABILITY DISTRIBUTION

When two or more random variables X and Y are studied simultaneously on a sample space, we get a joint probability distribution. Consider the experiment of throwing two unbiased dice. If X denotes the number on the first and Y denotes the number on the second die, then X and Y are random variables having a joint probability distribution. When the number of random variables is two, it is called a *bi-variate probability distribution* and if the number of random variables become more than two, the distribution is termed as a *multivariate probability distribution*.

Let the random variable X take values X_1, X_2, \dots, X_m and Y take values Y_1, Y_2, \dots, Y_n . Further, let p_{ij} be the joint probability that X takes the value X_i and Y takes the value Y_j , i.e., $P[X = X_i \text{ and } Y = Y_j] = p_{ij}$ ($i = 1$ to m and $j = 1$ to n). This bi-variate probability distribution can be written in a tabular form as follows :

	Y_1	Y_2	Y_n	<i>Marginal Probabilities of X</i>
X_1	p_{11}	p_{12}	p_{1n}	P_1
X_2	p_{21}	p_{22}	p_{2n}	P_2
.
X_m	p_{m1}	p_{m2}	p_{mn}	P_m
<i>Marginal Probabilities of Y</i>	P'_1	P'_2	P'_n	1

12.5.1 Marginal Probability Distribution

In the above table, the probabilities given in each row are added and shown in the last column. Similarly, the sum of probabilities of each column are shown in the last row of the table. These probabilities are termed as marginal probabilities. The last column of the table gives the *marginal probabilities* for various values of random variable X. The set of all possible values of the random variable X along with their respective marginal probabilities is termed as the marginal probability distribution of X. Similarly, the marginal probabilities of the random variable Y are given in the last row of the above table.

Remarks: If X and Y are independent random variables, then by multiplication theorem of probability we have

$$P(X = X_i \text{ and } Y = Y_j) = P(X = X_i).P(Y = Y_j) \quad \forall \ i \text{ and } j$$

Using notations, we can write $p_{ij} = P_i.P'_j$

The above relation is similar to the relation between the relative frequencies of independent attributes.

12.5.2 Conditional Probability Distribution

Each column of the above table gives the probabilities for various values of the random variable X for a given value of Y, represented by it. For example, column 1 of the table represents that $P(X_1, Y_1) = p_{11}$, $P(X_2, Y_1) = p_{21}$, ..., $P(X_m, Y_1) = p_{m1}$, where $P(X_i, Y_1) = p_{i1}$ denote the probability of the event that $X = X_i$ ($i = 1$ to m) and $Y = Y_1$. From the conditional probability theorem, we can write

$$P(X = X_i / Y = Y_1) = \frac{\text{Joint probability of } X_i \text{ and } Y_1}{\text{Marginal probability of } Y_1} = \frac{p_{ij}}{P'_j} \quad (\text{for } i = 1, 2, \dots, m).$$

This gives us a conditional probability distribution of X given that $Y = Y_1$. This distribution can be written in a tabular form as shown below :

X	X_1	X_2	X_m	Total Probability
Probability	$\frac{p_{11}}{P'_1}$	$\frac{p_{21}}{P'_1}$	$\frac{p_{m1}}{P'_1}$	1

The conditional distribution of X given some other value of Y can be constructed in a similar way. Further, we can construct the conditional distributions of Y for various given values of X.

Remarks: It can be shown that if the conditional distribution of a random variable is same as its marginal distribution, the two random variables are independent. Thus, if for

the conditional distribution of X given Y_1 we have $\frac{p_{i1}}{P'_1} = P_i$ for $\forall i$, then X and Y are

independent. It should be noted here that if one conditional distribution satisfies the condition of independence of the random variables, then all the conditional distributions would also satisfy this condition.

Example 9: Let two unbiased dice be tossed. Let a random variable X take the value 1 if first die shows 1 or 2, value 2 if first die shows 3 or 4 and value 3 if first die shows 5 or 6. Further, Let Y be a random variable which denotes the number obtained on the second die. Construct a joint probability distribution of X and Y. Also determine their marginal probability distributions and find $E(X)$ and $E(Y)$ respectively. Determine the conditional distribution of X given $Y = 5$ and of Y given $X = 2$. Find the expected values of these conditional distributions. Determine whether X and Y are independent?

Solution: For the given random experiment, the random variable X takes values 1, 2 and 3 and the random variable Y takes values 1, 2, 3, 4, 5 and 6. Their joint probability distribution is shown in the following table :

$X \downarrow \setminus Y \rightarrow$	1	2	3	4	5	6	Marginal Dist. of X
1	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
2	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
3	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
Marginal Dist. of Y	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

From the above table, we can write the marginal distribution of X as given below :

X	1	2	3	Total
P_i	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Thus, the expected value of X is $E(X) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} = 2$

Similarly, the probability distribution of Y is

Y	1	2	3	4	5	6	Total
P'_j	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

and $E(Y) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5$

The conditional distribution of X when Y = 5 is

X	1	2	3	Total
$P_i / Y = 5$	$\frac{1}{18} \times \frac{6}{1} = \frac{1}{3}$	$\frac{1}{18} \times \frac{6}{1} = \frac{1}{3}$	$\frac{1}{18} \times \frac{6}{1} = \frac{1}{3}$	1

$$\therefore E(X / Y = 5) = \frac{1}{3}(1 + 2 + 3) = 2$$

The conditional distribution of Y when X = 2 is

Y	1	2	3	4	5	6	Total
$P'_j / X = 2$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$$\therefore E(Y / X = 2) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

Since the conditional distribution of X is same as its marginal distribution (or equivalently the conditional distribution of Y is same as its marginal distribution), X and Y are independent random variables.

Example 10: Two unbiased coins are tossed. Let X be a random variable which denotes the total number of heads obtained on a toss and Y be a random variable which takes a value 1 if head occurs on first coin and takes a value 0 if tail occurs on it. Construct the joint probability distribution of X and Y. Find the conditional distribution of X when Y = 0. Are X and Y independent random variables?

Solution: There are 4 elements in the sample space of the random experiment. The possible values that X can take are 0, 1 and 2 and the possible values of Y are 0 and 1. The joint probability distribution of X and Y can be written in a tabular form as follows :

$X \downarrow \setminus Y \rightarrow$	0	1	Total
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
Total	$\frac{2}{4}$	$\frac{2}{4}$	1

The conditional distribution of X when Y = 0, is given by

X	0	1	2	Total
$P(X / Y = 0)$	$\frac{1}{2}$	$\frac{1}{2}$	0	1

Also, the marginal distribution of X, is given by

X	0	1	2	Total
P_i	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Since the conditional and the marginal distributions are different, X and Y are not independent random variables.

12.5.3 Expectation of the Sum or Product of two Random Variables

Theorem 1: If X and Y are two random variables, then $E(X + Y) = E(X) + E(Y)$.

Proof: Let the random variable X takes values X_1, X_2, \dots, X_m and the random variable Y takes values Y_1, Y_2, \dots, Y_n such that $P(X = X_i \text{ and } Y = Y_j) = p_{ij}$ ($i = 1$ to m , $j = 1$ to n).

By definition of expectation, we can write

$$\begin{aligned} E(X + Y) &= \sum_{i=1}^m \sum_{j=1}^n (X_i + Y_j) p_{ij} = \sum_{i=1}^m \sum_{j=1}^n X_i p_{ij} + \sum_{i=1}^m \sum_{j=1}^n Y_j p_{ij} \\ &= \sum_{i=1}^m X_i \sum_{j=1}^n p_{ij} + \sum_{i=1}^n Y_j \sum_{j=1}^m p_{ij} \\ &= \sum_{i=1}^m X_i P_i + \sum_{j=1}^n Y_j P'_j \quad \left(\text{Here } \sum_{j=1}^n p_{ij} = P_i \text{ and } \sum_{i=1}^m p_{ij} = P'_j \right) \\ &= E(X) + E(Y) \end{aligned}$$

The above result can be generalised. If there are k random variables X_1, X_2, \dots, X_k , then $E(X_1 + X_2 + \dots + X_k) = E(X_1) + E(X_2) + \dots + E(X_k)$.

Remarks: The above result holds irrespective of whether X_1, X_2, \dots, X_k are independent or not.

Theorem 2: If X and Y are two independent random variables, then

$$E(XY) = E(X)E(Y)$$

Proof: Let the random variable X takes values X_1, X_2, \dots, X_m and the random variable Y takes values Y_1, Y_2, \dots, Y_n such that $P(X = X_i \text{ and } Y = Y_j) = p_{ij}$ ($i = 1$ to m , $j = 1$ to n).

$$\text{By definition } E(XY) = \sum_{i=1}^m \sum_{j=1}^n X_i Y_j p_{ij}$$

Since X and Y are independent, we have $p_{ij} = P_i P'_j$

$$\begin{aligned} \therefore E(XY) &= \sum_{i=1}^m \sum_{j=1}^n X_i Y_j P_i P'_j = \sum_{i=1}^m X_i P_i \times \sum_{j=1}^n Y_j P'_j \\ &= E(X)E(Y). \end{aligned}$$

The above result can be generalised. If there are k independent random variables X_1, X_2, \dots, X_k , then

$$E(X_1 \cdot X_2 \cdot \dots \cdot X_k) = E(X_1)E(X_2) \cdot \dots \cdot E(X_k)$$

12.5.4 Expectation of a Function of Random Variables

Let $\phi(X, Y)$ be a function of two random variables X and Y. Then we can write

$$E[\phi(X, Y)] = \sum_{i=1}^m \sum_{j=1}^n \phi(X_i, Y_j) p_{ij}$$

I. Expression for Covariance

As a particular case, assume that $\phi(X_i, Y_j) = (X_i - \mu_X)(Y_j - \mu_Y)$, where $E(X) = \mu_X$ and $E(Y) = \mu_Y$

$$\text{Thus, } E[(X - \mu_X)(Y - \mu_Y)] = \sum_{i=1}^m \sum_{j=1}^n (X_i - \mu_X)(Y_j - \mu_Y) p_{ij}$$

The above expression, which is the mean of the product of deviations of values from their respective means, is known as the Covariance of X and Y denoted as $\text{Cov}(X, Y)$ or σ_{XY} . Thus, we can write

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

An alternative expression of $\text{Cov}(X, Y)$

$$\begin{aligned}\text{Cov}(X, Y) &= E[\{X - E(X)\}\{Y - E(Y)\}] \\ &= E[X.\{Y - E(Y)\} - E(X).\{Y - E(Y)\}] \\ &= E[X.Y - X.E(Y)] = E(X.Y) - E(X).E(Y)\end{aligned}$$

Note that $E[\{Y - E(Y)\}] = 0$, the sum of deviations of values from their arithmetic mean.

Remarks: If X and Y are independent random variables, the right hand side of the above equation will be zero. Thus, covariance between independent variables is always equal to zero.

II. Mean and Variance of a Linear Combination

Let $Z = \phi(X, Y) = aX + bY$ be a linear combination of the two random variables X and Y, then using the theorem of addition of expectation, we can write

$$\mu_Z = E(Z) = E(aX + bY) = aE(X) + bE(Y) = a\mu_X + b\mu_Y$$

Further, the variance of Z is given by

$$\begin{aligned}\sigma_Z^2 &= E[Z - E(Z)]^2 = E[aX + bY - a\mu_X - b\mu_Y]^2 = E[a(X - \mu_X) + b(Y - \mu_Y)]^2 \\ &= a^2 E(X - \mu_X)^2 + b^2 E(Y - \mu_Y)^2 + 2abE(X - \mu_X)(Y - \mu_Y) \\ &= a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2ab\sigma_{XY}\end{aligned}$$

Remarks:

1. The above results indicate that any function of random variables is also a random variable.
2. If X and Y are independent, then $\sigma_{XY} = 0$, $\therefore \sigma_Z^2 = a^2 \sigma_X^2 + b^2 \sigma_Y^2$
3. If $Z = aX - bY$, then we can write $\sigma_Z^2 = a^2 \sigma_X^2 + b^2 \sigma_Y^2 - 2ab\sigma_{XY}$. However, $\sigma_Z^2 = a^2 \sigma_X^2 + b^2 \sigma_Y^2$, if X and Y are independent.
4. The above results can be generalised. If X_1, X_2, \dots, X_k are k independent random variables with means $\mu_1, \mu_2, \dots, \mu_k$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ respectively, then

$$E(X_1 \pm X_2 \pm \dots \pm X_k) = \mu_1 \pm \mu_2 \pm \dots \pm \mu_k$$

$$\text{and } \text{Var}(X_1 \pm X_2 \pm \dots \pm X_k) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2$$

Notes:

1. The general result on expectation of the sum or difference will hold even if the random variables are not independent.
2. The above result can also be proved for continuous random variables.

Example 11: A random variable X has the following probability distribution :

X	:	2	1	0	1	2
<i>Probability</i>	:	$\frac{1}{6}$	p	$\frac{1}{4}$	p	$\frac{1}{6}$

- (i) Find the value of p.
- (ii) Calculate $E(X + 2)$ and $E(2X^2 + 3X + 5)$.

Solution: Since the total probability under a probability distribution is equal to unity, the

value of p should be such that $\frac{1}{6} + p + \frac{1}{4} + p + \frac{1}{6} = 1$.

This condition gives $p = \frac{5}{24}$

Further, $E(X) = -2 \cdot \frac{1}{6} - 1 \cdot \frac{5}{24} + 0 \cdot \frac{1}{4} + 1 \cdot \frac{5}{24} + 2 \cdot \frac{1}{6} = 0$,

$E(X^2) = 4 \cdot \frac{1}{6} + 1 \cdot \frac{5}{24} + 0 \cdot \frac{1}{4} + 1 \cdot \frac{5}{24} + 4 \cdot \frac{1}{6} = \frac{7}{4}$,

$E(X + 2) = E(X) + 2 = 0 + 2 = 2$

and $E(2X^2 + 3X + 5) = 2E(X^2) + 3E(X) + 5 = 2 \cdot \frac{7}{4} + 0 + 5 = 8.5$

Example 12: A dealer of ceiling fans has estimated the following probability distribution of the price of a ceiling fan in the next summer season :

<i>Price (P)</i>	:	800	825	850	875	900
<i>Probability (p)</i>	:	0.15	0.25	0.30	0.20	0.10

If the demand (x) of his ceiling fans follows a linear relation $x = 6000 - 4P$, find expected demand of fans and expected total revenue of the dealer.

Solution: Since P is a random variable, therefore, $x = 6000 - 4P$, is also a random variable. Further, Total Revenue $TR = P.x = 6000P - 4P^2$ is also a random variable.

From the given probability distribution, we have

$$E(P) = 800 \times 0.15 + 825 \times 0.25 + 850 \times 0.30 + 875 \times 0.20 + 900 \times 0.10 \\ = \text{Rs } 846.25 \text{ and}$$

$$E(P^2) = (800)^2 \times 0.15 + (825)^2 \times 0.25 + (850)^2 \times 0.30 + (875)^2 \times 0.20 + (900)^2 \times 0.10 = 717031.25$$

Thus, $E(X) = 6000 - 4E(P) = 6000 - 4 \times 846.25 = 2615$ fans.

$$\text{And } E(TR) = 6000E(P) - 4E(P^2) \\ = 6000 \times 846.25 - 4 \times 717031.25 = \text{Rs } 22,09,375.00$$

Example 13: A person applies for equity shares of Rs 10 each to be issued at a premium of Rs 6 per share; Rs 8 per share being payable along with the application and the balance at the time of allotment. The issuing company may issue 50 or 100 shares to those who apply for 200 shares, the probability of issuing 50 shares being 0.4 and that of issuing 100 shares is 0.6. In either case, the probability of an application being selected for allotment of any shares is 0.2. The allotment usually takes three months and the market price per share is expected to be Rs 25 at the time of allotment. Find the expected rate of return of the person per month.

Solution: Let A be the event that the application of the person is considered for allotment, B_1 be the event that he is allotted 50 shares and B_2 be the event that he is allotted 100 shares. Further, let R_1 denote the rate of return (per month) when 50 shares are allotted, R_2 be the rate of return when 100 shares are allotted and $R = R_1 + R_2$ be the combined rate of return.

We are given that $P(A) = 0.2$, $P(B_1/A) = 0.4$ and $P(B_2/A) = 0.6$.

(a) **When 50 shares are allotted**

The return on investment in 3 months = $(25 - 16)50 = 450$

$$\therefore \text{Monthly rate of return} = \frac{450}{3} = 150$$

The probability that he is allotted 50 shares

$$= P(A \cap B_1) = P(A).P(B_1 / A) = 0.2 \times 0.4 = 0.08$$

Thus, the random variable R_1 takes a value 150 with probability 0.08 and it takes a value 0 with probability $1 - 0.08 = 0.92$

$$\therefore E(R_1) = 150 \times 0.08 + 0 = 12.00$$

(b) **When 100 shares are allotted**

The return on investment in 3 months = $(25 - 16).100 = 900$

$$\therefore \text{Monthly rate of return} = \frac{900}{3} = 300$$

The probability that he is allotted 100 shares

$$= P(A \cap B_2) = P(A).P(B_2 / A) = 0.2 \times 0.6 = 0.12$$

Thus, the random variable R_2 takes a value 300 with probability 0.12 and it takes a value 0 with probability $1 - 0.12 = 0.88$

$$\therefore E(R_2) = 300 \times 0.12 + 0 = 36$$

$$\text{Hence, } E(R) = E(R_1 + R_2) = E(R_1) + E(R_2) = 12 + 36 = 48$$

Example 14: What is the mathematical expectation of the sum of points on n unbiased dice?

Solution: Let X_i denote the number obtained on the i th die. Therefore, the sum of points on n dice is $S = X_1 + X_2 + \dots + X_n$ and

$$E(S) = E(X_1) + E(X_2) + \dots + E(X_n).$$

Further, the number on the i th die, i.e., X_i follows the following distribution :

$$\begin{array}{rccccc} X_i & : & 1 & 2 & 3 & 4 & 5 & 6 \\ p(X_i) & : & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \therefore E(X_i) & = & \frac{1}{6}(1+2+3+4+5+6) & = & \frac{7}{2} & (i = 1, 2, \dots, n) \end{array}$$

$$\text{Thus, } E(S) = \frac{7}{2} + \frac{7}{2} + \dots + \frac{7}{2} \text{ (n times)} = \frac{7n}{2}$$

Example 15: If X and Y are two independent random variables with means 50 and 120 and variances 10 and 12 respectively, find the mean and variance of $Z = 4X + 3Y$.

$$\text{Solution: } E(Z) = E(4X + 3Y) = 4E(X) + 3E(Y) = 4 \times 50 + 3 \times 120 = 560$$

Since X and Y are independent, we can write

$$\text{Var}(Z) = \text{Var}(4X + 3Y) = 16\text{Var}(X) + 9\text{Var}(Y) = 16 \times 10 + 9 \times 12 = 268$$

Example 16: It costs Rs 600 to test a machine. If a defective machine is installed, it costs Rs 12,000 to repair the damage resulting to the machine. Is it more profitable to install the machine without testing if it is known that 3% of all the machines produced are defective? Show by calculations.

Solution: Here X is a random variable which takes a value 12,000 with probability 0.03 and a value 0 with probability 0.97.

$$\therefore E(X) = 12000 \times 0.03 + 0 \times 0.97 = \text{Rs } 360.$$

Since $E(X)$ is less than Rs 600, the cost of testing the machine, hence, it is more profitable to install the machine without testing.

Exercise with Hints

- A man draws two balls at random from a bag containing three white and five black balls. If he is to receive Rs 14 for every white ball that he draws and Rs 7 for every black ball, what should be his expectation of earning in the game?

Hint: Random variable takes 3 values 14, 21 and 28.

- ABC company estimates the net profit on a new product, that it is launching, to be Rs 30,00,000 if it is successful, Rs 10,00,000 if it is moderately successful and a loss of Rs 10,00,000 if it is unsuccessful. The firm assigns the following probabilities to the different possibilities : Successful 0.15, moderately successful 0.25 and unsuccessful 0.60. Find the expected value and variance of the net profits.

Hint: See example 5.

- There are 4 different choices available to a customer who wants to buy a transistor set. The first type costs Rs 800, the second type Rs 680, the third type Rs 880 and the fourth type Rs 760. The probabilities that the customer will buy these types are

$\frac{1}{3}, \frac{1}{6}, \frac{1}{4}$ and $\frac{1}{4}$ respectively. The retailer of these sets gets a commission @ 20%,

12%, 25% and 15% on the respective sets. What is the expected commission of the retailer?

Hint: Take commission as random variable.

- Three cards are drawn at random successively, with replacement, from a well shuffled pack of cards. Getting a card of diamond is termed as a success. Tabulate the probability distribution of the number successes (X). Find the mean and variance of X.

Hint: The random variable takes values 0, 1, 2 and 3.

- A discrete random variable can take all possible integral values from 1 to k each with probability $\frac{1}{k}$. Find the mean and variance of the distribution.

$$\text{Hint: } E(X^2) = \frac{1}{k} (1^2 + 2^2 + \dots + k^2) = \frac{1}{k} \left[\frac{k(k+1)(2k+1)}{6} \right].$$

- An insurance company charges, from a man aged 50, an annual premium of Rs 15 on a policy of Rs 1,000. If the death rate is 6 per thousand per year for this age group, what is the expected gain for the insurance company?

Hint: Random variable takes values 15 and - 985.

- On buying a ticket, a player is allowed to toss three fair coins. He is paid number of rupees equal to the number of heads appearing. What is the maximum amount the player should be willing to pay for the ticket.

Hint: The maximum amount is equal to expected value.

- The following is the probability distribution of the monthly demand of calculators :

Demand (x) : 15 16 17 18 19 20

Probability p(x) : 0.10 0.15 0.35 0.25 0.08 0.07

Calculate the expected demand for calculators. If the cost c of producing x calculators is given by the relation $c = 4x^2 - 15x + 200$, find expected cost.

Hint: See example 12.

9. Firm A wishes to bid for the supply of 800 chairs to an educational institution at the rate of Rs 500 per chair. The firm, which has two competitors B and C, has estimated that the probability that firm B will bid less than Rs 500 per chair is 0.4 and that the firm C will bid less than Rs 500 per chair is 0.6. If the lowest bidder gets business and the firms bid independently, what is the expected value of the contract to firm A?

Hint: The random variable takes value 0 with probability 0.4×0.6 and it takes value 500×800 with probability $1 - 0.4 \times 0.6$.

10. A game is played by throwing a six faced die for which the incomplete probability distribution of the number obtained is given below :

$$\begin{array}{ccccccc} X & : & 1 & 2 & 3 & 4 & 5 & 6 \\ p(X) & : & 0.09 & 0.30 & m & n & 0.28 & 0.09 \end{array}$$

The conditions of the game are : If the die shows an even number, the player gets rupees equal to the number obtained; if the die shows 3 or 5, he loses rupees equal to the number obtained, while if 1 is obtained the player neither gains or loses. Complete the probability distribution if the game is given to be fair.

Hint: $E(X) = 0$ for a fair game.

11. There are three bags which contain 4 red and 3 black, 6 red and 4 black and 8 red and 2 black balls respectively. One ball is drawn from each urn. What is the expected number of red balls obtained?

Hint: Find the expected number of red balls from each urn and add.

12. A survey conducted over last 25 years indicated that in 10 years the winter was mild, in 8 years it was cold and in the remaining 7 years it was very cold. A company sells 1,000 woollen coats in mild cold year, 1,300 in a cold year and 2,000 in a very cold year.

You are required to find the yearly expected profit of the company if a woollen coat costs Rs 173 and is sold to stores for Rs 248.

Hint: The random variable can take 3 possible values.

13. You have been offered the chance to play a dice game in which you will receive Rs 20 each time the point total of a toss of two dice is 6. If it costs you Rs 2.50 per toss to participate, should you play or not? Will it make any difference in your decision if it costs Rs 3.00 per toss instead of Rs 2.50?

Hint: Compare the cost of participation with the expected value of the receipt.

14. The probability that a house of a certain type will be on fire in a year is 0.005. An insurance company offers to sell the owner of such a house Rs 1,00,000 one year term insurance policy for a premium of Rs 600. What is the expected gain of the company?

Hint: See exercise 6.

15. Three persons A, B and C in that order draw a ball, without replacement, from a bag containing 2 red and 3 white balls till someone is able to draw a red ball. One who draws a red ball wins Rs 400. Determine their expectations.

Hint: A wins if he gets a red ball on the first draw or all the three get white ball in their respective first draws, etc.

16. A coin is tossed until a head appears. What is the expected number and standard deviation of tosses?

Hint: The random variable takes values 1, 2, 3, ... with respective probabilities p , $(1 - p)p$, $(1 - p)^2p$, etc., where p is the probability of getting a head.

17. A box contains 8 tickets. 3 of the tickets carry a prize of Rs 5 each and the remaining 5 a prize of Rs 2 each.

- (i) If one ticket is drawn at random, what is the expected value of the prize?
- (ii) If two tickets are drawn at random, what is the expected value of the prize?

Hint: (i) The random variable can take values 5 or 2, (ii) It can take values 4, 7 or 10.

18. 4 unbiased coins are tossed 256 times. Find the frequency distribution of heads and tabulate the result. Calculate the mean and standard deviation of the number of heads.

Hint: the random variable takes values 0, 1, 2, 3 and 4.

19. Throwing two unbiased coins simultaneously, Mr X bets with Mrs X that he will receive Rs 4 from her if he gets 2 heads and he will give Rs 4 to her otherwise. Find Mr X's expectation.

Hint: The random variable takes values 4 and -4.

20. A man runs an ice cream parlor in a holiday resort. If the summer is mild, he can sell 2,500 cups of ice cream; if it is hot, he can sell 4,000 cups; if it is very hot, he can sell 5,000 cups. It is known that for any year the probability of the summer to be mild is $\frac{1}{7}$ and to be hot is $\frac{4}{7}$. A cup of ice cream costs Rs 2 and sold for Rs 3.50. What is his expected profit?

Hint: See example 5.

21. Comment on the validity of the following statement :

$$\text{For a random variable } X, \sqrt{E(X^2)} \geq E(X).$$

Hint: $\sigma^2 = E(X^2) - [E(X)]^2$.

Check Your Progress 12.1

1. What is Stochastic variable?
2. How bi-variate probability is different from multi variable probability distribution?

Notes: (a) Write your answer in the space given below.

- (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-
-

12.6 DECISION ANALYSIS UNDER CERTAINTY

Decision-making is needed whenever an individual or an organisation (private or public) is faced with a situation of selecting an optimal (or best in view of certain objectives) course of action from among several available alternatives. For example, an individual may have to decide whether to build a house or to purchase a flat or live in a rented accommodation; whether to join a service or to start own business; which company's car should be purchased, etc. Similarly, a business firm may have to decide the type of technique to be used in production, what is the most appropriate method of advertising its product, etc.

The decision analysis provides certain criteria for the selection of a course of action such that the objective of the decision-maker is satisfied. The course of action selected on the basis of such criteria is termed as the *optimal course of action*.

Every decision problem has four basic features, mentioned below:

1. **Alternative Courses of Action or Acts:** Every decision-maker is faced with a set of several alternative courses of action A_1, A_2, \dots, A_m and he has to select one of them in view of the objectives to be fulfilled.
2. **States of Nature:** The consequences of selection of a course of action are dependent upon certain factors that are beyond the control of the decision-maker. These factors are known as states of nature or events. It is assumed that the decision-maker is aware of the whole list of events S_1, S_2, \dots, S_n and exactly one of them is bound to occur. In other words, the events S_1, S_2, \dots, S_n are assumed to be mutually exclusive and collective exhaustive.
3. **Consequences:** The results or outcomes of selection of a particular course of action are termed as its consequences. The consequence, measured in quantitative or value terms, is called payoff of a course of action. It is assumed that the payoffs of various courses of action are known to the decision-maker.
4. **Decision Criterion:** Given the payoffs of various combinations of courses of action and the states of nature, the decision-maker has to select an optimal course of action. The criterion for such a selection, however, depends upon the attitude of the decision-maker.

If X_{ij} denotes the payoff corresponding to a combination of a course of action and a state of nature, i.e., (A_i, S_j) , $i = 1$ to m and $j = 1$ to n , the above elements of a decision problem can be presented in a matrix form, popularly known as the Payoff Matrix.

Payoff Matrix

<i>Events → Actions ↓</i>	S_1	S_2	...	S_j	...	S_n
A_1	X_{11}	X_{12}	...	X_{1j}	...	X_{1n}
A_2	X_{21}	X_{22}	...	X_{2j}	...	X_{2n}
⋮	⋮	⋮		⋮		⋮
A_i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{in}
⋮	⋮	⋮		⋮		⋮
A_m	X_{m1}	X_{m2}	...	X_{mj}	...	X_{mn}

Given the payoff matrix for a decision problem, the process of decision-making depends upon the situation under which the decision is being made. These situations can be classified into three broad categories : (a) Decision-making under certainty, (b) Decision-making under uncertainty and (c) Decision-making under risk.

Decision-making under Certainty

Probability Distribution
of a Random Variable

The conditions of certainty are very rare particularly when significant decisions are involved. Under conditions of certainty, the decision-maker knows which particular state of nature will occur or equivalently, he is aware of the consequences of each course of action with certainty. Under such a situation, the decision-maker should focus on the corresponding column in the payoff table and choose a course of action with optimal payoff.

12.7 DECISION-MAKING UNDER UNCERTAINTY

A situation of uncertainty arises when there can be more than one possible consequences of selecting any course of action. In terms of the payoff matrix, if the decision-maker selects A_1 , his payoff can be X_{11} , X_{12} , X_{13} , etc., depending upon which state of nature S_1 , S_2 , S_3 , etc., is going to occur. A decision problem, where a decision-maker is aware of various possible states of nature but has insufficient information to assign any probabilities of occurrence to them, is termed as decision-making under uncertainty.

There are a variety of criteria that have been proposed for the selection of an optimal course of action under the environment of uncertainty. Each of these criteria make an assumption about the attitude of the decision-maker.

1. **Maximin Criterion:** This criterion, also known as the criterion of pessimism, is used when the decision-maker is pessimistic about future. Maximin implies the maximisation of minimum payoff. The pessimistic decision-maker locates the minimum payoff for each possible course of action. The maximum of these minimum payoffs is identified and the corresponding course of action is selected. This is explained in the following example :

Example 17: Let there be a situation in which a decision-maker has three possible alternatives A_1 , A_2 and A_3 , where the outcome of each of them can be affected by the occurrence of any one of the four possible events S_1 , S_2 , S_3 and S_4 . The monetary payoffs of each combination of A_i and S_j are given in the following table :

Payoff Matrix

<i>Events →</i> <i>Actions ↓</i>	S_1	S_2	S_3	S_4	<i>Min. Payoff</i>	<i>Max. Payoff</i>
A_1	27	12	14	26	12	27
A_2	45	17	35	20	17	45
A_3	52	36	29	15	15	52

Solution: Since 17 is maximum out of the minimum payoffs, the optimal action is A_2 .

2. **Maximax Criterion:** This criterion, also known as the criterion of optimism, is used when the decision-maker is optimistic about future. Maximax implies the maximisation of maximum payoff. The optimistic decision-maker locates the maximum payoff for each possible course of action. The maximum of these payoffs is identified and the corresponding course of action is selected. The optimal course of action in the above example, based on this criterion, is A_3 .
3. **Regret Criterion:** This criterion focuses upon the regret that the decision-maker might have from selecting a particular course of action. Regret is defined as the difference between the best payoff we could have realised, had we known which state of nature was going to occur and the realised payoff. This difference, which measures the magnitude of the loss incurred by not selecting the best alternative, is also known as opportunity loss or the *opportunity cost*.

From the payoff matrix (given in § 12.6), the payoffs corresponding to the actions A_1, A_2, \dots, A_n under the state of nature S_j are $X_{1j}, X_{2j}, \dots, X_{nj}$ respectively. Of these assume that X_{2j} is maximum. Then the regret in selecting A_i to be denoted by R_{ij} is given by $X_{2j} - X_{ij}$, $i = 1$ to m . We note that the regret in selecting A_2 is zero. The regrets for various actions under different states of nature can also be computed in a similar way.

The regret criterion is based upon the minimax principle, i.e., the decision-maker tries to minimise the maximum regret. Thus, the decision-maker selects the maximum regret for each of the actions and out of these the action which corresponds to the minimum regret is regarded as optimal.

The regret matrix of example 17 can be written as given below:

Regret Matrix

<i>Events → Actions ↓</i>	S_1	S_2	S_3	S_4	<i>Max. Regret</i>
A_1	25	24	21	0	25
A_2	7	19	0	6	19
A_3	0	0	6	11	11

From the maximum regret column, we find that the regret corresponding to the course of action is A_3 is minimum. Hence, A_3 is optimal.

4. **Hurwicz Criterion:** The maximax and the maximin criteria, discussed above, assumes that the decision-maker is either optimistic or pessimistic. A more realistic approach would, however, be to take into account the degree or *index of optimism* or *pessimism* of the decision-maker in the process of decision-making. If a , a constant lying between 0 and 1, denotes the degree of optimism, then the degree of pessimism will be $1 - a$. Then a weighted average of the maximum and minimum payoffs of an action, with a and $1 - a$ as respective weights, is computed. The action with highest average is regarded as optimal.

We note that a nearer to unity indicates that the decision-maker is optimistic while a value nearer to zero indicates that he is pessimistic. If $a = 0.5$, the decision-maker is said to be neutralist.

We apply this criterion to the payoff matrix of example 17. Assume that the index of optimism $a = 0.7$.

Action	Max. Payoff	Min. Payoff	Weighted Average
A_1	27	12	$27 \times 0.7 + 12 \times 0.3 = 22.5$
A_2	45	17	$45 \times 0.7 + 17 \times 0.3 = 36.6$
A_3	52	15	$52 \times 0.7 + 15 \times 0.3 = 40.9$

Since the average for A_3 is maximum, it is optimal.

5. **Laplace Criterion:** In the absence of any knowledge about the probabilities of occurrence of various states of nature, one possible way out is to assume that all of them are equally likely to occur. Thus, if there are n states of nature, each can be assigned a probability of occurrence = $1/n$. Using these probabilities, we compute the expected payoff for each course of action and the action with maximum expected value is regarded as optimal.

12.8 DECISION-MAKING UNDER RISK

Probability Distribution
of a Random Variable

In case of decision-making under uncertainty the probabilities of occurrence of various states of nature are not known. When these probabilities are known or can be estimated, the choice of an optimal action, based on these probabilities, is termed as decision-making under risk.

The choice of an optimal action is based on The Bayesian Decision Criterion according to which an action with maximum Expected Monetary Value (EMV) or minimum Expected Opportunity Loss (EOL) or Regret is regarded as optimal.

Example 18: The payoffs (in Rs) of three Acts A_1 , A_2 and A_3 and the possible states of nature S_1 , S_2 and S_3 are given below :

		<i>Acts →</i>	A_1	A_2	A_3
		<i>States of Nature ↓</i>			
		S_1	-20	-50	200
		S_2	200	-100	-50
		S_3	400	600	300

The probabilities of the states of nature are 0.3, 0.4 and 0.3 respectively. Determine the optimal act using the Bayesian Criterion.

Solution:

Computation of Expected Monetary Value

	S_1	S_2	S_3	
$P(S)$	0.3	0.4	0.3	EMV
A_1	-20	200	400	$-20 \times 0.3 + 200 \times 0.4 + 400 \times 0.3 = 194$
A_2	-50	-100	600	$-50 \times 0.3 - 100 \times 0.4 + 600 \times 0.3 = 125$
A_3	200	-50	300	$200 \times 0.3 - 50 \times 0.4 + 300 \times 0.3 = 130$

From the above table, we find that the act A_1 is optimal.

The problem can alternatively be attempted by finding minimum EOL, as shown below:

Computation of Expected Opportunity Loss

	S_1	S_2	S_3	
$P(S)$	0.3	0.4	0.3	EOL
A_1	220	0	200	$220 \times 0.3 + 0 \times 0.4 + 200 \times 0.3 = 126$
A_2	250	300	0	$250 \times 0.3 + 300 \times 0.4 + 0 \times 0.3 = 195$
A_3	0	250	300	$0 \times 0.3 + 250 \times 0.4 + 300 \times 0.3 = 190$

This indicates that the optimal act is again A_1 .

12.9 EXPECTED VALUE WITH PERFECT INFORMATION (EVPI)

The expected value with perfect information is the amount of profit foregone due to uncertain conditions affecting the selection of a course of action.

Given the perfect information, a decision-maker is supposed to know which particular state of nature will be in effect. Thus, the procedure for the selection of an optimal course of action, for the decision problem given in example 18, will be as follows :

If the decision-maker is certain that the state of nature S_1 will be in effect, he would select the course of action A_3 , having maximum payoff equal to Rs 200.

Similarly, if the decision-maker is certain that the state of nature S_2 will be in effect, his course of action would be A_1 and if he is certain that the state of nature S_3 will be in effect, his course of action would be A_2 . The maximum payoffs associated with the actions are Rs 200 and Rs 600 respectively.

The weighted average of these payoffs with weights equal to the probabilities of respective states of nature is termed as *Expected Payoff under Certainty* (EPC).

$$\text{Thus, } EPC = 200 \times 0.3 + 200 \times 0.4 + 600 \times 0.3 = 320$$

The difference between EPC and EMV of optimal action is the amount of profit foregone due to uncertainty and is equal to EVPI.

$$\text{Thus, } EVPI = EPC - EMV \text{ of optimal action} = 320 - 194 = 126$$

It is interesting to note that EVPI is also equal to EOL of the optimal action.

Cost of Uncertainty

This concept is similar to the concept of EVPI. Cost of uncertainty is the difference between the EOL of optimal action and the EOL under perfect information.

Given the perfect information, the decision-maker would select an action with minimum opportunity loss under each state of nature. Since minimum opportunity loss under each state of nature is zero, therefore,

$$\text{EOL under certainty} = 0 \times 0.3 + 0 \times 0.4 + 0 \times 0.3 = 0 .$$

Thus, the cost of uncertainty = EOL of optimal action = EVPI

Example 19: A group of students raise money each year by selling souvenirs outside the stadium of a cricket match between teams A and B. They can buy any of three different types of souvenirs from a supplier. Their sales are mostly dependent on which team wins the match. A conditional payoff (in Rs.) table is as under :

Type of Souvenir →	I	II	III
Team A wins	1200	800	300
Team B wins	250	700	1100

- (i) Construct the opportunity loss table.
- (ii) Which type of souvenir should the students buy if the probability of team A's winning is 0.6?
- (iii) Compute the cost of uncertainty.

Solution:

- (i) **The Opportunity Loss Table**

Actions →	Type of Souvenir bought		
	I	II	III
Events ↓			
Team A wins	0	400	900
Team B wins	850	400	0

$$(ii) \text{ EOL of buying type I Souvenir} = 0 \times 0.6 + 850 \times 0.4 = 340$$

$$\text{EOL of buying type II Souvenir} = 400 \times 0.6 + 400 \times 0.4 = 400.$$

$$\text{EOL of buying type III Souvenir} = 900 \times 0.6 + 0 \times 0.4 = 540.$$

Since the EOL of buying Type I Souvenir is minimum, the optimal decision is to buy Type I Souvenir.

- (iii) Cost of uncertainty = EOL of optimal action = Rs. 340

Example 20: The following is the information concerning a product X :

Probability Distribution
of a Random Variable

- (i) Per unit profit is Rs 3.
- (ii) Salvage loss per unit is Rs 2.
- (iii) Demand recorded over 300 days is as under :

<i>Units demanded</i>	: 5 6 7 8 9
<i>No. of days</i>	: 30 60 90 75 45

Find : (i) EMV of optimal order.

(ii) Expected profit presuming certainty of demand.

Solution:

- (i) The given data can be rewritten in terms of relative frequencies, as shown below:

<i>Units demanded</i>	: 5 6 7 8 9
<i>No. of days</i>	: 0.1 0.2 0.3 0.25 0.15

From the above probability distribution, it is obvious that the optimum order would lie between and including 5 to 9.

Let A denote the number of units ordered and D denote the number of units demanded per day.

If $D \geq A$, profit per day = $3A$, and if $D < A$, profit per day = $3D - 2(A - D) = 5D - 2A$.

Thus, the profit matrix can be written as

<i>Units Demanded</i>	5	6	7	8	9	
<i>Probability →</i>	0.10	0.20	0.30	0.25	0.15	<i>EMV</i>
<i>Action (units ordered) ↓</i>						
5	15	15	15	15	15	15.00
6	13	18	18	18	18	17.50
7	11	16	21	21	21	19.00
8	9	14	19	24	24	19.00
9	7	12	17	22	27	17.75

From the above table, we note that the maximum EMV = 19.00, which corresponds to the order of 7 or 8 units. Since the order of the 8th unit adds nothing to the EMV, i.e., marginal EMV is zero, therefore, order of 8 units per day is optimal.

- (ii) Expected profit under certainty

$$= (5 \times 0.10 + 6 \times 0.20 + 7 \times 0.30 + 8 \times 0.25 + 9 \times 0.15) \times 3 = \text{Rs } 21.45$$

Alternative Method: The work of computations of EMV's, in the above example, can be reduced considerably by the use of the concept of expected marginal profit. Let π be the marginal profit and λ be the marginal loss of ordering an additional unit of the product. Then, the expected marginal profit of ordering the A th unit, is given by

$$= \pi \cdot P(D \geq A) - \lambda \cdot P(D < A) = \pi \cdot P(D \geq A) - \lambda \cdot [1 - P(D \geq A)]$$

$$= (\pi + \lambda) \cdot P(D \geq A) - \lambda \quad \dots (1)$$

The computations of EMV, for alternative possible values of A, are shown in the following table :

In our example, $\pi = 3$ and $\lambda = 2$

Thus, the expression for the expected marginal profit of the Ath unit

$$= (3+2)P(D \geq A) - 2 = 5P(D \geq A) - 2.$$

Table for Computations

Action(A)	$P(D \geq A)^*$	$EMP = 5P(D \geq A) - 2$	Total profit or EMV
5	1.00	$5 \times 1.00 - 2 = 3.00$	$5 \times 3.00 = 15.00$
6	0.90	$5 \times 0.90 - 2 = 2.50$	$15.00 + 2.50 = 17.50$
7	0.70	$5 \times 0.70 - 2 = 1.50$	$17.50 + 1.50 = 19.00$
8	0.40	$5 \times 0.40 - 2 = 0.00$	$19.00 + 0.00 = 19.00$
9	0.15	$5 \times 0.15 - 2 = -1.25$	$19.00 - 1.25 = 17.75$

* This column represents the 'more than type' cumulative probabilities.

Since the expected marginal profit (EMP) of the 8th unit is zero, therefore, optimal order is 8 units.

Marginal Analysis

Marginal analysis is used when the number of states of nature is considerably large. Using this analysis, it is possible to locate the optimal course of action without the computation of EMV's of various actions.

An order of A units is said to be optimal if the expected marginal profit of the Ath unit is non-negative and the expected marginal profit of the (A + 1)th unit is negative. Using equation (1), we can write

$$(\pi + \lambda)P(D \geq A) - \lambda \geq 0 \quad \text{and} \quad \dots (2)$$

$$(\pi + \lambda)P(D \geq A + 1) - \lambda < 0 \quad \dots (3)$$

From equation (2), we get

$$P(D \geq A) \geq \frac{\lambda}{\pi + \lambda} \text{ or } 1 - P(D < A) \geq \frac{\lambda}{\pi + \lambda}$$

$$\text{or } P(D < A) \leq 1 - \frac{\lambda}{\pi + \lambda} \text{ or } P(D \leq A - 1) \leq \frac{\pi}{\pi + \lambda} \quad \dots (4)$$

[$P(D \leq A - 1) = P(D < A)$, since A is an integer]

Further, equation (3) gives

$$P(D \geq A + 1) < \frac{\lambda}{\pi + \lambda} \text{ or } 1 - P(D < A + 1) < \frac{\lambda}{\pi + \lambda}$$

$$\text{or } P(D < A + 1) > 1 - \frac{\lambda}{\pi + \lambda} \text{ or } P(D \leq A) > \frac{\pi}{\pi + \lambda} \quad \dots (5)$$

Combining (4) and (5), we get

$$P(D \leq A - 1) \leq \frac{\pi}{\pi + \lambda} < P(D \leq A).$$

Writing the probability distribution, given in example 20, in the form of less than type cumulative probabilities which is also known as the distribution function F(D), we get

<i>Units demanded(D)</i>	: 5 6 7 8 9
<i>F(D)</i>	: 0.1 0.3 0.6 0.85 1.00

$$\text{We are given } \pi = 3 \text{ and } \lambda = 2, \therefore \frac{\pi}{\pi + \lambda} = \frac{3}{5} = 0.6$$

Since the next cumulative probability, i.e., 0.85, corresponds to 8 units, hence, the optimal order is 8 units.

12.10 USE OF SUBJECTIVE PROBABILITIES IN DECISION-MAKING

When the objective probabilities of the occurrence of various states of nature are not known, the same can be assigned on the basis of the expectations or the degree of belief of the decision-maker. Such probabilities are known as subjective or personal probabilities. It may be pointed out that different individuals may assign different probability values to given states of nature.

This indicates that a decision problem under uncertainty can always be converted into a decision problem under risk by the use of subjective probabilities. Such an approach is also termed as *Subjectivists' Approach*.

Example 21: The conditional payoff (in Rs) for each action-event combination are as under:

Action →	1	2	3	4
Event ↓	1	2	3	4
A	4	-2	7	8
B	0	6	3	5
C	-5	9	2	-3
D	3	1	4	5
E	6	6	3	2

- (i) Which is the best action in accordance with the Maximin Criterion?
- (ii) Which is the best action in accordance with the EMV Criterion, assuming that all the events are equally likely?

Solution:

- (i) The minimum payoffs for various actions are :

Action 1 = -5

Action 2 = -2

Action 3 = 2

Action 4 = -3

Since the payoff for action 3 is maximum, therefore, A₃ is optimal on the basis of maximin criterion.

- (ii) Since there are 5 equally likely events, the probability of each of them would be $\frac{1}{5}$.

$$\text{Thus, the EMV of action 1, i.e., } EMV_1 = \frac{4+0-5+3+6}{5} = \frac{8}{5} = 1.6$$

$$\text{Similarly, } EMV_2 = \frac{20}{5} = 4.0, EMV_3 = \frac{19}{5} = 3.8 \text{ and } EMV_4 = \frac{17}{5} = 3.4$$

Thus, action 2 is optimal.

12.11 USE OF POSTERIOR PROBABILITIES IN DECISION-MAKING

The probability values of various states of nature, discussed so far, were prior probabilities. Such probabilities are either computed from the past data or assigned subjectively. It is possible to revise these probabilities in the light of current information available by using the Bayes' Theorem. The revised probabilities are known as posterior probabilities.

Example 22: A manufacturer of detergent soap must determine whether or not to expand his productive capacity. His profit per month, however, depend upon the potential demand for his product which may turn out to be high or low. His payoff matrix is given below:

	Do not Expand	Expand
High Demand	Rs 5,000	Rs 7,500
Low Demand	Rs 5,000	Rs 2,100

On the basis of past experience, he has estimated the probability that demand for his product being high in future is only 0.4

Before taking a decision, he also conducts a market survey. From the past experience he knows that when the demand has been high, such a survey had predicted it correctly only 60% of the times and when the demand has been low, the survey predicted it correctly only 80% of the times.

If the current survey predicts that the demand of his product is going to be high in future, determine whether the manufacturer should increase his production capacity or not? What would have been his decision in the absence of survey?

Solution: Let H be the event that the demand will be high. Therefore,

$$P(H) = 0.4 \text{ and } P(\bar{H}) = 0.6$$

Note that H and \bar{H} are the only two states of nature.

Let D be the event that the survey predicts high demand. Therefore,

$$P(D/H) = 0.60 \text{ and } P(\bar{D}/\bar{H}) = 0.80$$

We have to find $P(H/D)$ and $P(\bar{H}/D)$. For this, we make the following table:

	H	\bar{H}	Total
D	$0.4 \times 0.6 = 0.24$	0.12	0.36
\bar{D}	0.16	$0.6 \times 0.8 = 0.48$	0.64
Total	0.40	0.60	1.00

From the above table, we can write

$$P(H/D) = \frac{0.24}{0.36} = \frac{2}{3} \text{ and } P(\bar{H}/D) = \frac{0.12}{0.36} = \frac{1}{3}$$

The EMV of the act 'don't expand' = $5000 \times \frac{2}{3} + 5000 \times \frac{1}{3} = \text{Rs } 5,000$

and the EMV of the act 'expand' = $7500 \times \frac{2}{3} + 2100 \times \frac{1}{3} = \text{Rs } 5,700$

Since the EMV of the act 'expand' > the EMV of the act 'don't expand', the manufacturer should expand his production capacity.

It can be shown that, in the absence of survey the EMV of the act 'don't expand' is Rs 5,000 and the EMV of the act expand is Rs 4,260. Hence, the optimal act is 'don't expand'.

Decision Tree Approach

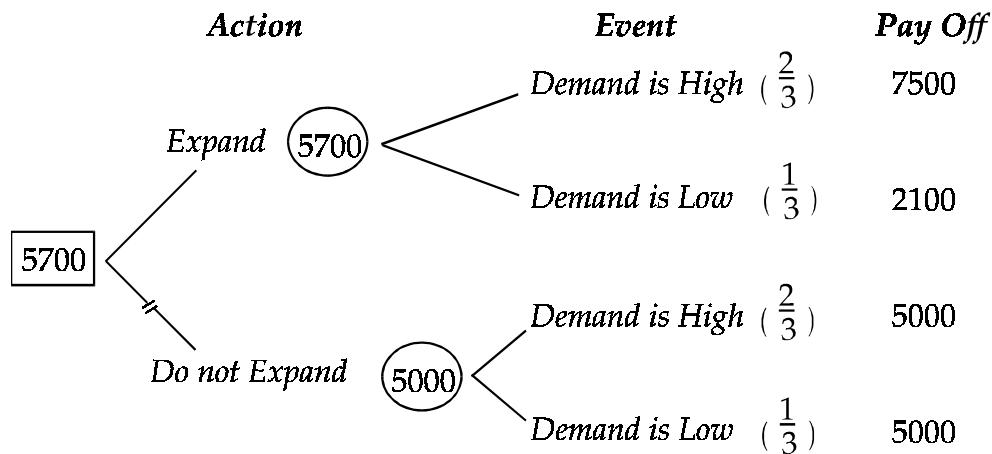
Probability Distribution
of a Random Variable

The decision tree diagrams are often used to understand and solve a decision problem. Using such diagrams, it is possible to describe the sequence of actions and chance events. A decision node is represented by a square and various action branches stem from it. Similarly, a chance node is represented by a circle and various event branches stem from it. Various steps in the construction of a decision tree can be summarised as follows :

- (i) Show the appropriate action-event sequence beginning from left to right of the page.
- (ii) Write the probabilities of various events along their respective branches stemming from each chance node.
- (iii) Write the payoffs at the end of each of the right-most branch.
- (iv) Moving backward, from right to left, compute EMV of each chance node, wherever encountered. Enter this EMV in the chance node. When a decision node is encountered, choose the action branch having the highest EMV. Enter this EMV in the decision node and cutoff the other action branches.

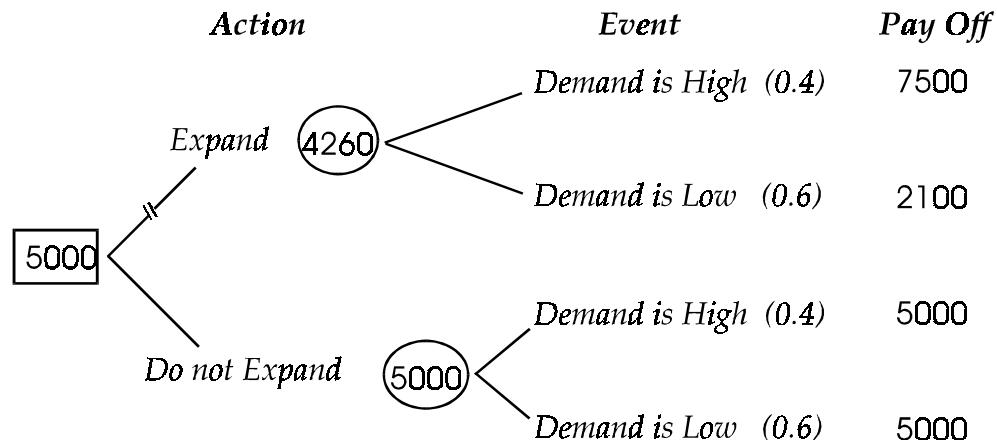
Following this approach, we can describe the decision problem of the above example as given below:

Case I: When the survey predicts that the demand is going to be high



Thus, the optimal act to expand capacity.

Case II: In the absence of survey



Thus, the optimal act is not to expand capacity.

Exercise with Hints

- The probability of the demand for lorries for hire on any day in a given district is as follows :

<i>No. of lorries demanded</i>	: 0	1	2	3	4
<i>Probability</i>	: 0.1	0.2	0.3	0.2	0.2

Lorries have a fixed cost of Rs 90 each day to keep and the daily hire charge (net of variable costs of running) is 200. If the lorry-hire company owns 4 lorries, what is its daily expectations? If the lorry-hire company is about to go into business and currently has no lorries, how many lorries should it buy?

Hint: Take $\pi = 110$ and $\lambda = 90$.

- A management is faced with the problem of choosing one of the products for manufacturing. The potential demand for each product may turn out to be good, moderate or poor. The probabilities for each of the states of nature were estimated as follows :

<i>Nature of Demand →</i>		<i>Good</i>	<i>Moderate</i>	<i>Poor</i>
<i>Product ↓</i>				
X		0.70	0.20	0.10
Y		0.50	0.30	0.20
Z		0.40	0.50	0.10

The profit or loss (in Rs) under the three states is estimated as

X	30,000	20,000	10,000
Y	60,000	30,000	20,000
Z	40,000	10,000	-15,000

Prepare the expected value table and advise the management about the choice of product.

Hint: Compute expected profit for each commodity.

- A pig breeder can either produce 20 or 30 pigs. The total production of his competitors can be either 5,000 or 10,000 pigs. If they produce 5,000 pigs, his profit per pig is Rs 60; if they produce 10,000 pigs, his profit per pig is Rs 45 only. Construct a payoff table and also state what should the pig breeder decide?

Hint: This is a decision problem under uncertainty where the courses of actions are to produce 20 or 30 pigs while the states of nature are the production of 5,000 or 10,000 pigs by his competitors.

- Mr X quite often flies from town A to town B. He can use the airport bus which costs Rs 13 but if he takes it, there is a 0.08 chance that he will miss the flight. A hotel limousine costs Rs. 27 with a 0.96 chance of being on time for the flight. For Rs 50 he can use a taxi which will make 99 of 100 flights. If Mr X catches the flight on time, he will conclude a business transaction which will produce a profit of Rs 1,000; otherwise he will lose it. Which mode of transportation should Mr X use? Answer on the basis of EMV criterion.

Hint: EMV of using airport bus = $(1000 - 13) \times 0.92 - 13 \times 0.08$, etc.

- A distributor of a certain product incurs holding cost of Rs 100 per unit per week and a shortage cost of Rs 300 per unit. The data on the sales of the product are given below :

<i>Weekly Sales</i>	:	0	1	2	3	4	5	6	7	8
<i>No. of Weeks</i>	:	0	0	5	10	15	15	5	0	0

Find his optimal stock.

Hint: Take $\pi = 300$ and $\lambda = 100$.

Check Your Progress 12.2

Probability Distribution
of a Random Variable

- 1 Distinguish between Hurwicz Criterion is different from Laplace Criterion.
2. What is the use of subjective and posterior probabilities in decision-making?

Notes: (a) Write your answer in the space given below.

- (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
- (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-

12.12 LET US SUM UP

In this chapter you would be able to understand probability distribution of a random variable and also mean and variance. In this various theorems on expectation to find out expected value. And taking decisions on the basis of various probabilistic distribution.

1. Mean of a discrete random variable is $E(X) = \sum_{i=1}^n X_i \cdot p(X_i)$
2. $\text{Var}(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$
3. $E(b) = b$, where b is a constant
4. $E(aX + b) = aE(X) + b$
5. $\text{Var}(aX + b) = a^2\text{Var}(X)$
6. $E(X + Y) = E(X) + E(Y)$
7. $E(XY) = E(X)E(Y)$, if X and Y are independent.
8. Bayesian Decision Criterion : An action with maximum EMV or minimum EOL is said to be optimal.

12.13 LESSON-END ACTIVITY

Use the probability distribution in a laboratory for experiments.

12.14 KEYWORDS

Variable

Decision Analysis

Variance

Theorems

Marginal Analysis

12.15 QUESTIONS FOR DISCUSSION

1. Fill in the blanks:

- (a) Random variable expressed in monetary units, its expected value is
- (b) Simultaneously studying two or more random variables is called probability distribution.
- (c) If X and Y are two random variables, then $= E(X) + E(Y)$
- (d) A situation of uncertainty arises when there can be more than one consequence of any course of action.
- (e) is the amount of profit foregone due to uncertainty.

2. Write True or False against each statement:

- (a) Cost of uncertainty is the difference between EOL of optimal and perfect information.
- (b) In the toss of B coins assuming that they are unbiased the probability is $1/8$.
- (c) Marginal Analysis is used when no. of states of nature is small.
- (d) Subjective and personal probabilities are same.
- (e) Decision tree is used to solve decision problem.

3. Write short notes on:

- (a) Pay-off Matrix
- (b) Theorems on Expectation
- (c) Decision Analysis
- (d) Joint Probability Distribution
- (e) Action and Event

4. Distinguish Between:

- (a) Action and Event
- (b) Discrete and Continuous Probability Distribution
- (c) Bivariate and Multivariate Probability Distribution
- (d) Joint and Marginal Probability
- (e) EMV and EOL

12.16 TERMINAL QUESTIONS

1. Explain the concept of random variable and its probability distribution by using a simple example.
2. What is mathematical expectation of a random variable? If $Y = aX + b$, where X is a random variable, show that $E(Y) = aE(X) + b$.
3. If X and Y are two independent random variables, show that
 - (a) $E(X + Y) = E(X) + E(Y)$
 - (b) $E(X.Y) = E(X).E(Y)$
4. A bag contains 3 rupee coins, 6 fifty paise coins and 4 twenty-five paise coins. A man draws a coin at random. What is the expectation of his draw?
5. A box contains five tickets; two of which carry a prize of Rs 8 each and the other three of Rs 3 each. If two tickets are drawn at random, find the expected value of the prize.

6. Obtain the probability distribution of the number of aces in simultaneous throws of two unbiased dice.
7. You are told that the time to service a car at a service station is uncertain with following probability density function:

$$f(x) = 3x - 2x^2 + 1 \text{ for } 0 \leq x \leq 2$$

$$= 0 \text{ otherwise.}$$

Examine whether this is a valid probability density function?

8. Find mean and variance of the following probability distribution :

X	: 20	10	30
$p(X)$: $\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{2}$

9. An urn contains 4 white and 3 black balls. 3 balls are drawn at random. Write down the probability distribution of the number of white balls. Find mean and variance of the distribution.
10. A consignment is offered to two firms A and B for Rs 50,000. The following table shows the probability at which the firm will be able to sell it at different prices :

Selling Price (in Rs)	40,000	45,000	55,000	70,000
Prob. of A	0.3	0.4	0.2	0.1
Prob. of B	0.1	0.2	0.4	0.3

Which of the two firms will be more inclined towards the offer?

11. If the probability that the value of a certain stock will remain same is 0.46, the probabilities that its value will increase by Re. 0.50 or Re. 1.00 per share are respectively 0.17 and 0.23 and the probability that its value will decrease by Re. 0.25 per share is 0.14, what is the expected gain per share?
12. In a college fete a stall is run where on buying a ticket a person is allowed one throw of two dice. If this gives a double six, 10 times the ticket money is refunded and in other cases nothing is refunded. Will it be profitable to run such a stall? What is the expectation of the player? State clearly the assumptions if any, for your answer.
13. The proprietor of a food stall has introduced a new item of food. The cost of making it is Rs 4 per piece and because of its novelty, it would be sold for Rs 8 per piece. It is, however, perishable and pieces remaining unsold at the end of the day are a dead loss. He expects the daily demand to be variable and has drawn up the following probability distribution expressing his estimates:

No. of pieces demanded	:	50	51	52	53	54	55
Probability	:	0.05	0.07	0.20	0.35	0.25	0.08

Compute his expected profit or loss if he prepares 53 pieces on a particular day.

14. The probability that there is at least one error in an accounts statement prepared by A is 0.2 and for B and C are 0.25 and 0.4 respectively. A, B and C prepare 10, 16 and 20 statements respectively. Find the expected number of correct statements in all.
15. Three coins whose faces are marked as 1 and 2 are tossed. What is the expectation of the total value of numbers on their faces?
16. A person has the choice of running hot snack stall or an ice cream and cold drink shop at a certain holiday resort during the coming summer season. If the weather during the season is cool and rainy, he can expect to make a profit of Rs 15,000 and if it is warm, he can expect to make a profit of Rs 3,000 only, by running a hot snack stall. On the other hand, if his choice is to run an ice cream and cold drink

shop, he can expect to make a profit of Rs 18,000 if the weather is warm and only Rs 3,000 if the weather is cool and rainy. The meteorological authorities predict that there is 40% chance of the weather being warm during the coming season. You are to advise him as to the choice between the two types of stalls. Base your argument on the expectation of the result of the two courses of action and show the result in a tabular form.

17. Show that the expectation of the number of failures preceding the first success in an infinite series of independent trials is q/p , where p is the probability of success in a single trial and $q = 1 - p$.
18. If X is a random variable with expected value 50 and standard deviation 4, find the values of a and b such that the expected value of $Y = aX + b$ is zero and standard deviation is 6.
19. A discrete random variable X has the following probability distribution:

X	:	0	1	2	3	4	5
$p(X)$:	k	$2k$	$3k$	$5k$	$4k$	$3k$

Find (a) the value of k , (b) $P(X \geq 3)$, (c) the value of m such that $P(X \leq m) = \frac{5}{6}$ and (d) write the distribution function of X .

20. A company introduces a new product in the market and expects to make a profit of Rs 2.5 lacs during first year if the demand is 'good', Rs 1.5 lacs if the demand is 'moderate' and a loss of Rs 1 lac if the demand is 'poor'. Market research studies indicate that the probabilities for the demand to be good and moderate are 0.2 and 0.5 respectively. Find the company's expected profit and standard deviation.
21. If it rains, a taxi driver can earn Rs 100 per day. If it is fair, he can lose Rs 10 per day. What is his expectation if the probability of rain is 0.4?
22. A player tosses 3 fair coins. He wins Rs 10 if three heads appear, Rs 6 if two heads appear, Rs 2 if one head appears and loses Rs 25 if no head appears. Find the expected gain of the player.
23. A player tosses 3 fair coins. He wins Rs 12 if three tails occur, Rs 7 if two tails occur and Rs 2 if only one tail occurs. How much should he win or lose in case of occurrence of no tail if the game is given to be fair?
24. A firm plans to bid Rs 300 per tonne for a contract to supply 1,000 tonnes of a metal. It has two competitors A and B and it assumes that the probability that A will bid less than Rs 300 per tonne is 0.3 and that B will bid less than Rs 300 per tonne is 0.7. If the lowest bidder gets all the business and the firms bid independently, what is the expected value of the contract to the firm?

25. A certain production process produces items that are 10 percent defective. Each item is inspected before being supplied to customers but the inspector incorrectly classifies an item 10 percent of the times. Only items classified as good are supplied. If 820 items in all have been supplied, how many of these are expected to be defective?

Hint: Let A be the event that an item is supplied. $P(A) = 0.10 \times 0.10 + 0.90 \times 0.90 = 0.82$. Let B be the event that a defective item is supplied. $P(B) = 0.10 \times 0.10 = 0.01$. Therefore $P(B/A) = 0.01/0.82$.

26. You are given the following payoffs of three acts A_1 , A_2 and A_3 and the states of nature S_1 , S_2 and S_3 :

States of Nature	Acts		
	A_1	A_2	A_3
S_1	25	10	125
S_2	400	440	400
S_3	650	740	750

The probabilities of the three states of nature are 0.1, 0.7 and 0.2 respectively. Compute and tabulate the EMV and determine the optimal act.

27. Given is the following payoff (in Rs) matrix :

State of Nature	Probability	Decision		
		Do not	Expand	Expand
		Expand	200 units	400 units
High Demand	0.4	2500	3500	5000
Medium Demand	0.4	2500	3500	2500
Low Demand	0.2	2500	1500	1000

What should be the decision if we use (i) EMV criterion, (ii) The minimax criterion and (iii) the maximin criterion?

28. The proprietor of a food stall has invented a new food delicacy which he calls WHIM. He has calculated that the cost of manufacture is Re 1 per piece and because of its novelty, it can be sold for Rs 3 per piece. It is, however, perishable and the goods unsold at the end of the day are a dead loss. He expects the demand to be variable and has drawn up the following probability distribution of his estimate:

No. of pieces demanded	:	10	11	12	13	14	15
Probability	:	0.07	0.10	0.23	0.38	0.12	0.10

- (i) Find an expression for his net profit or loss if he manufacture m pieces and only n are demanded. Consider separately the two cases $n \leq m$ and $n > m$.
- (ii) Assume that he manufactures 12 pieces. Using the results in (i) above, find his net profit or loss for each level of demand.
- (iii) Using the probability distribution, calculate his expected net profit or loss if he manufactures 12 pieces.
- (iv) Calculate the expected profit or loss for each of the levels of manufacture ($10 \leq m \leq 15$).
- (v) How many pieces should be manufactured so that his expected profit is maximum?

29. A physician purchases a particular vaccine on Monday of each week. The vaccine must be used in the current week, otherwise it becomes worthless. The vaccine costs Rs 2 per dose and the physician charges Rs 4 per dose. In the past 50 weeks, the physician has administered the vaccine in the following quantities :

Doses per week	:	20	25	40	60
No. of weeks	:	5	15	25	5

Determine the number of doses the physician should buy every week.

30. The marketing staff of a certain industrial organisation has submitted the following payoff table, giving profits in million rupees, concerning a proposal depending upon the rate of technological advance in the next three years :

Technological advance	Accept	Reject
	Proposal	Proposal
Much	2	3
Little	5	2
None	1	4

The probabilities are 0.2, 0.5 and 0.3 for Much, Little and None technological advance respectively. What decision should be taken?

31. A newspaper distributor assigns probabilities to the demand for a magazine as follows:

<i>Copies Demanded</i>	: 1	2	3	4
<i>Probability</i>	: 0.4	0.3	0.2	0.1

A copy of magazine sells for Rs 7 and costs Rs 6. What can be the maximum possible expected monetary value (EMV) if the distributor can return the unsold copies for Rs 5 each? Also find EVPI.

32. A management is faced with the problem of choosing one of the three products for manufacturing. The potential demand for each product may turn out to be good, fair or poor. The probabilities for each type of demand were estimated as follows:

<i>Demand →</i>		<i>Good</i>	<i>Fair</i>	<i>Poor</i>
<i>Product ↓</i>				
	<i>A</i>	0.75	0.15	0.10
	<i>B</i>	0.60	0.30	0.10
	<i>C</i>	0.50	0.30	0.20

The estimated profit or loss (in Rs) under the three states of demand in respect of each product may be taken as :

<i>A</i>	35,000	15,000	5,000
<i>B</i>	50,000	20,000	3,000
<i>C</i>	60,000	30,000	20,000

Prepare the expected value table and advise the management about the choice of the product.

33. The payoffs of three acts A, B and C and the states of nature P, Q and R are given as :

<i>States of Nature</i>	<i>Payoffs (in Rs)</i>		
	<i>A</i>	<i>B</i>	<i>C</i>
<i>P</i>	-35	120	-100
<i>Q</i>	250	-350	200
<i>R</i>	550	650	700

The probabilities of the states of nature are 0.5, 0.1 and 0.4 respectively. Tabulate the Expected Monetary Values for the above data and state which can be chosen as the best act? Calculate expected value of perfect information also.

34. A manufacturing company is faced with the problem of choosing from four products to manufacture. The potential demand for each product may turn out to be good, satisfactory or poor. The probabilities estimated of each type of demand are given below :

<i>Product</i>	<i>Probabilities of type of demand</i>		
	<i>Good</i>	<i>Satisfactory</i>	<i>Poor</i>
<i>A</i>	0.60	0.20	0.20
<i>B</i>	0.75	0.15	0.10
<i>C</i>	0.60	0.25	0.15
<i>D</i>	0.50	0.20	0.30

The estimated profit (in Rs) under different states of demand in respect of each product may be taken as :

<i>A</i>	40,000	10,000	1,100
<i>B</i>	40,000	20,000	7,000
<i>C</i>	50,000	15,000	8,000
<i>D</i>	40,000	18,000	15,000

Prepare the expected value table and advise the company about the choice of product to manufacture.

35. A shopkeeper at a local stadium must determine whether to sell ice cream or coffee at today's game. The shopkeeper believes that the profit will depend upon the weather. The payoff table is as follows :

Event	Action	
	Sell Coffee	Sell Ice cream
Cool Weather	Rs 40	Rs 20
Warm Weather	Rs 55	Rs 80

Based upon his past experience at this time of the year, the shopkeeper estimates the probability of warm weather as 0.60. Prior to making his decision, the shopkeeper decides to hear forecast of the local weatherman. In the past, when it has been cool, the weatherman has forecast cool weather 80% times. When it has been warm, the weatherman has forecast warm weather 70% times. If today's forecast is for cool weather, using Bayesian decision theory and EMV criterion, determine whether the shopkeeper should sell ice cream or coffee?

36. A producer of boats has estimated the following distribution of demand for a particular kind of boat :

Demand	:	0	1	2	3	4	5	6
Probability	:	0.14	0.27	0.27	0.18	0.09	0.04	0.01

Each boat costs him Rs 7,000 and he sells them for Rs 10,000 each. Any boats that are left unsold at the end of the season must be disposed off for Rs 6,000 each. How many boats should be kept in stock to maximise his expected profit?

37. A retailer purchases berries every morning at Rs 5 a case and sells for Rs 8 a case. Any case remaining unsold at the end of the day can be disposed of the next day at a salvage value of Rs 2 per case (thereafter they have no value). Past sales have ranged from 15 to 18 cases per day. The following is the record of sales for the past 120 days :

No. of cases sold	:	15	16	17	18
No. of days	:	12	24	48	36

Find how many cases the retailer should purchase per day to maximise his profit?

38. State whether the following statements are True or False :

- (i) A random variable takes a value corresponding to every element of the sample space.
- (ii) The probability of a given value of the discrete random variable is obtained by its probability density function.
- (iii) Distribution function is another name of cumulative probability function.
- (iv) Any function of a random variable is also a random variable.
- (v) The expected value of the sum of two or more random variables is equal to the sum of their expected values only if the are independent.
- (vi) In the process of decision-making, the decision-maker can also assign probabilities to various states of nature based upon his degree of belief.

39. **Fill in blanks :**

- (i) The probability that a random variable takes a particular value is always zero.
- (ii) The mean of a random variable is also termed as its value.
- (iii) Any function of random variable is also a

- (iv) If the conditional distribution of X given Y is same as the marginal distribution of X, then X and Y are random variables.
- (v) The selection of a particular decision criterion depends upon the of the decision-maker.
- (vi) An action with maximum EMV or minimum EOL is regarded as

12.17 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

1. (a) Expected Monetary Value (EMV) (b) Joint (c) $\in (X + Y)$
(d) Possible (e) Expected Value with Perfect Information (EVPI)
2. (a) True (b) True (c) False (d) True (e) True

12.18 SUGGESTED READINGS

Lerry Gonick and Woollcott Smith, *The Cartoon Guide to Statistics*, Harpercollins, New York 1994.

Bernard W. Lindgren, *Statistical Theory*, Macmillan, third edition 1976.

Siegried Graf, Harold Luschgy, *Foundations of quantizations for probability distributions*, Springer-Verlag Berlin Heidelberg New York 1993

Edgar H. Callaway, David A. Hill, Jerome V. Scholle, *Probability Distribution*, Motorola Univ. April 1993

Gejza Wimmes, Gabriel Altmann, *Thesaurus of Universe Discrete Probability Distribution*, Stany, Jan. 1999

Unit-V

LESSON

13

INVENTORY MODEL

CONTENTS

- 13.0 Aims and Objectives
- 13.1 Introduction
- 13.2 Need of Inventory Control
- 13.3 Advantages of Material Controls
- 13.4 Essential Factors of Material Control
- 13.5 ABC Analysis Technique
- 13.6 Process of Inventory Control
- 13.7 Minimum Stock Level
- 13.8 Maximum Stock Level
- 13.9 Ordering Level or Re-order Level
- 13.10 Average Stock level
- 13.11 Danger Level
- 13.12 Let us Sum Up
- 13.13 Lesson-end Activities
- 13.14 Keywords
- 13.15 Questions for Discussion
- 13.16 Terminal Questions
- 13.17 Model Answers to Questions for Discussion
- 13.18 Suggested Readings

13.0 AIMS AND OBJECTIVES

In this lesson we are going to talk about the inventory control and how to minimise the total cost of inventory. The inventory control system offers comprehensive report capabilities to keep you on top of inventory status. It can facilitate in bringing about the creating of new or improved purchasing policies, save policies, pricing modelling and even enhanced customer service.

13.1 INTRODUCTION

The inventory means a physical stocks of good which is kept in hand for smooth and efficient running of future affairs of an organisation at the minimum costs of funds blocked in inventories. In a manufacturing organisation, inventory control plays a significant role because the total investment in inventories of various kinds is quite substantial. In this chapter we are going to discuss the meaning of inventory, need to control inventory, advantage of material control, essential factor of material control, the ABC analysis techniques, process of inventory control.

Inventory can be defined as the stock of goods, commodities or other resources that are stored at any given period for future production. In real, inventory control is a process itself, with the help of which, the demand of items, scheduling, purchase receiving, inspection, storage and despatch are arranged in such a manner that at minimum cost and in minimum time, the goods can be despatched to production department. Inventory control makes use of available capital in a most effective way and ensures adequate supply of goods for production.

13.2 NEED OF INVENTORY CONTROL

The main objectives of Inventory Control are as follows:

1. ***For Effective Cost Accounting System:*** Cost accounting system is useful only when there is a tight control over cost and inventory cost is a major part of total production cost.
2. ***To Check Waste and Wastage:*** Inventory control not just only ensures uninterrupted material supply to production department but also ensures the control from purchasing to supply of finished goods to customers. So in this way it checks waste and wastage whether it is about time, money or material.
3. ***To Check Embezzlement and Theft:*** Inventory control is to maintain necessary records for protecting theft and embezzlement.
4. ***For the Success of Business:*** Customer's satisfaction is very much important for the success of business and customer's satisfaction is directly related to the goods supplied to them. If the goods supplied to customers are low in cost with good quality at right time, it ensures the success of business. Inventory Control helps in achieving this goal.
5. ***For the Life of the Business:*** In absence of Inventory Control there are many risks of losses.
6. ***To Check National Wastage:*** Inventory control checks the wastage of nation's resources such as raw minerals, ores, etc.

13.3 ADVANTAGES OF MATERIAL CONTROLS

They are as follows:

1. It helps to minimise loss by obsolescence, deterioration damage etc.
2. It helps to protect against thefts, wastages, etc.
3. It helps managers in decision making.
4. To minimise capital investment in inventory.
5. To minimise cost of material purchasing.
6. To increase the storing capacity.
7. To maintain reasonable stocks of materials.
8. To facilitates regular and timely supply to customers.
9. To ensures smooth production operations.
10. To check national wastage.

13.4 ESSENTIAL FACTORS OF MATERIAL CONTROL

For the success of material control following factors should be kept in mind.

1. ***Proper Co-ordination:*** There should be a proper co-ordination between all the departments who uses materials, such as purchase department, store department, inspection department, accounts department, production department and sales department, so that there is neither a scarcity of material nor excess of material.

2. ***Centralisation of Purchasing:*** The important requirement of a successful inventory control system is the appointment of intelligent and experienced personnel in purchase department, these personnel should be expert in their field and negotiating the deals.
3. ***Proper Scheduling:*** All the requisitions made by production department should be scheduled, so material could be issued them by time and production should not be stopped.
4. ***Proper Classification:*** Classification and identification of inventories by allotting proper code number to each item and group should be done, to facilitate prompt recordings, locating and dealing.
5. ***Use of Standard Forms:*** Standards forms should be used so that any information can be send to all department within no time.
6. ***Internal Check System:*** Audit should be done by an independent party to check effectiveness of inventory control system.
7. ***Proper Storing System:*** Adequate and well organised warehouse facilities with well-equipped proper handling facilities must be there. Such facilities will reduce the wastage due to leakage, wear and tear, sustained dust and mishandling of materials. Store location should be in between the purchase department and production department, so that cost of internal transportation can be minimised.
8. ***Proper Store Accounting:*** An efficient inventory control necessitates maintenance of proper inventory records. Any typical information regarding any particular item of inventory may be taken from such records.
9. ***Proper Issuing System:*** There should be a well organised issuing system of material so that production process do not suffer.
10. ***Perpetual Inventory System:*** Daily stock position should be taken in this system.
11. ***Fixing of Various Stock Levels:*** Minimum stock level, maximum stock level, reorder point, safety level etc, should be pre-determined to ensure the continuity of smooth production.
12. ***Determination of Economic Order Quantity:*** Economic order quantity should be determined to minimise the cost of inventory.
13. ***Regular Reporting System:*** The information regarding the stock position, materials quantity etc, should be available to management regularly.

13.5 ABC ANALYSIS TECHNIQUE

Where there are a large number of items in the inventory, it becomes essential to have an efficient control over all items of stores. However comparatively, great care should be given to items of higher value. The movements of certain manufacturing concerns may consist of a small number of items representing a major portion of inventory value and a large number of items may represent a minor portion of inventory value. In such cases, a selective approach for inventory control should be followed.

The most modern technique for controlling the inventory is a value item analysis popularly known as ABC analysis which attempts to relate, how the inventory value is concentrated among the individual item. This analysis is based on Pareto's law. Pareto's law states that a few items of higher usage having high investment value should be paid more attention than a bulk of items having low usage value and having a low investment in capital. Under this analysis, all items of stores are divided into three main categories — A, B and C. Category A includes the most important items which represent about 60 to 70 per cent of the value of stores but constitute only 10 to 15 percent items. These items are recognised for special attention category B includes lesser important items representing an investment value of 20 to 25 percent and constitute a similar percentage of items of stores. Category C consists of the least important items of stores and constituted 60 to 70 percent of stores items representing only a capital investment between 10 to 15 percent. Close attention is paid to items falling in category A and the best items of category C. This classification of items into A, B and C categories is based upon value, usage, rate and criticality of items and these variables are given due weightage in categorising the items the term ABC implies Always Better Control.

Steps in ABC Analysis

Though no definite procedure can be laid down for classifying the inventories into A, B and C categories as this will depend upon a number of factors such as nature and varieties of items specific requirements of the business place of items in the production etc. These factors vary from business to business to business and items to item. However, following procedure can be followed:

- (i) First, the quality of each material expected to be used in a given period should be estimated.
- (ii) Secondly, the money value of the items of materials, so chosen should be calculated by multiplying the quantity of each item with the price.
- (iii) Thirdly, the items should be rearranged in the descending order of their value irrespective of their quantities.
- (iv) Fourthly, a running total of all the values and items will then be taken and then the figure so obtained should be converted into percentage of the gross total.
- (v) Fifthly and lastly, it will be found that a small number of a first few items may amount to a large percentage of the total value of the items. the management, then, will have to take a decision as to percentage of the total value or the total number of items which have to be covered by A, B and C categories.

Advantages of ABC Analysis

These are as follows:

1. **Increase in Profitability:** ABC analysis ensures a close control over the items of A, B and C categories and due to control over A category items, the capital investment over inventory reduces.
2. **Other Uses:** The technique of ABC analysis is based on the principle of management by exception and can be used in areas like, distribution, sales, etc.

13.6 PROCESS OF INVENTORY CONTROL

For the convenience to understand the topic, the inventory control system may be divided into three parts:

- (i) Process of Purchasing of Materials.
- (ii) Inventory Storing Procedure.
- (iii) Process of Issue of Materials.

Process of Purchasing of Materials

Its steps are as follows:

1. **Establishment of Purchase Department:** A different department should be established for purchase of materials. This department not only ensure the availability of raw material but also, machines, stationary etc. are purchased by this department.

Purchase of materials should be centralised. All purchase should be under a single department. Control centralised purchase is generally possible only in these industries, which are located at a single place only and nature of production is of same type. But if an industry has different production centre at different places, then it becomes compulsory to follow decentralised purchase system. Thus it is compulsory to have a complete knowledge about the nature of production, capacity of locality etc.

2. **Preparation of Purchasing Budget:** First of all the production target of the company should be determined, on the basis of which the budget for purchasing of material is prepared.

Following points should be kept in mind while preparing purchase budget:

- (i) System to receive the materials.
- (ii) The quantity and quality of the material according to the production requirements.

- (iii) Source of supply.
 - (iv) Present balance of materials and predictions to receive the materials ordered.
 - (v) Available cash for debtors.
 - (vi) On which date the indent is made by concerned department.
 - (vii) The conditions regarding the value of the material and rebate or discount on it.
3. ***Preparation of Purchase Requisition Slip:*** The initiations of purchase begins with the formal request from the various sections or departments to the purchase department to order goods. The request is made in a prescribed form to the purchase department by the departments needing the goods, authorising the purchase department for procuring the goods as per the specifications given in the slip by the date mentioned on it.

Specimen of a PRS

No. Pr Date:

Cost Centre

Katech Corporation Ltd Purchase Requisition Slip

Please purchase for department

Item No.	Code No.	Description	Quantity Required	Remark

Required by Checked by Approved by
.....

For use of department issuing this requisition			For use of Purchased department			
Item No.	Quantity in stock	Consumption per day/month	Quantity required	Purchase order no.	Supplier	Delivery Date

Store keeper

The requisitions are generally prepared in triplicate the original copy is sent to the purchase department, the second copy is retained by the store or the department initiating the purchase requisition and third are sent to the costing department.

4. ***Obtaining the Tender:*** After the decision for purchase tenders are invited from the prospective suppliers on studying the terms of supply and the quantity and quality of the goods. Vendor is selected out of the tenderers for the comparative study of tenderers. Following type of table may be used:

Type of Specimen of Tenderer Table

Katech Corporation Ltd.

Schedule of Quotations

Material Date S.No

Name of the party	Quantity offered	Rate/Unit	Terms	Time of delivery	Mode of delivery	Remarks

Store keeper Date

5. **Sending Purchase Order:** After comparing the difference tenderers, the best vendor is decided and the order of required material quotation is placed to him.

Purchase order is prepared in prescribed form by the purchase department and sent to the vendor authorising him to supply a specified quantity and quality of the materials at the stipulated terms at the time and place mentioned therein. Generally purchase order has the following information:

- (i) Name of the purchaser, serial no. and date of order.
- (ii) Name of vendor and address.
- (iii) Full details of materials quantity etc.
- (iv) Value, rebate and terms of payment etc.
- (v) Time and place of delivery.
- (vi) Directions regarding packing and despatching.
- (vii) Signature of purchaser.
- (viii) Method of follow-up.

Specimen of a Purchase Order

Katech Corporation Ltd.

Cable	S. No.
To,	Telephone
M/s	Date
.....	Reg. No.
.....	Our Ref.

Please supply the following items in accordance with the terms and conditions mentioned herein

Item No.	Description	Quantity	Price	Unit	Amount	Remarks

Terms and conditions:

Delivery at	Sales Tax
Discount	Freight
Excise Duty	Terms of Payment

For Katech Corporation Ltd.

(Signature)

Acknowledgement

Kindly acknowledge the receipt of this order:

Received on

Date of Delivery

Challan Date

Invoice No. Date

Specimen of Goods Received Note

Inventory Model

Katech Corporation Ltd.

Goods Received Note

From

M/s (Supplier) No. VRN

..... Date

.....

Goods		Quantity	No. of Packets etc.	Order No.	Delivery Note No.	Demanded by department	Remarks	Inspection	
Description	Code						Qty. rejected	Reason	
Carrier	Received by	Store A ledger	A/c & ref.	Inspector's signature					

6. **Receiving and Inspection of Materials:** When goods arrive they are taken delivery of and parcels or packet unpacked and the contents of the packages are checked by the receiving clerk with the order placed by the purchasing department to the vendor. After proper checking goods should be delivered to the laboratory or inspection department. Goods received note is prepared here.
7. **Returning the Materials:** On checking if any discrepancy is found as regards to quality and quantity. It should immediately be referred to the purchasing department so that the discrepancy may be adjusted or steps may be taken to return the defective or damaged goods in exchange of proper quality material on credit note.
8. **Payment of Purchased Material:** After required inspection etc. final report is sent to purchase officer, who sent it to payment officer after placing required entries in the report. After checking the ledger, payment officer authorise accounts clerk for payment.

Process of Issue of Materials

To control the issue of materials following procedure is followed:

1. **Issue of Materials:** When a foreman of any production department needs materials from store, he prepares three copies of goods requisition slip. If the material is costly and important then factory manager also sign these copies. One copy of requisition slip is kept by foreman itself and other two copies are given to stores. According to the requisition slip the store-keeper issues the materials to foreman. Foreman signs the two copy of store's requisition slips to verify that he has received the materials. then storekeeper makes the required entries in the bin card. After signing both the copies of requisition slip storekeeper sent one copy to accountant of store. After recording the issue of materials, store accountant sent this copy to costing department.

Sigma Corporation Ltd. Material Requisition Slip

To No

..... Date

Deliver following material to Fore order No.
and Job No.

Quantity	Unit	Description	Code No.	Office use only		Remarks
				Rate Amount	Issued by	
Authorised by		Store Ledger No.....		Issued by	Received by	Cost office Ref. No..... Price by

Sigma Corporation Ltd.

Materials Requisition Slip

Job No. Material Requisition Slip No.

Department Date

Please send the following materials.

Quantity	Code or Symbol	Description of Materials	Rate*	Amount*

Foreman Bin No.

Store Ledger Folio Returned

Storekeeper Cost Clerk

* Both these entries are to be done by cost clerk.

4. **Inter Departmental Transfer of Materials:** (For details see ‘Inventory Storing Procedure’)

ABC Co. Ltd.

Materials Transfer Slips

Issuing Department Serial No.

Receiving Department Date

Please receive the following materials.

Quantity	Code or Symbol	Description of Materials	Rate*	Amount	Reason for Transfer

..... Foreman Transfer Foreman Transferee Cost Clerk

* To be filled by Cost Clerk.

5. **To Prepare Material Abstract:** (For details see Inventory Storing System).

Sigma Corporation Ltd.

Material Abstract

Week ending on

Materials Requisition Slips		Job Numbers					
Slip No.	Amount	N ¹	N ²	N ³	N ⁴	N ⁵	N ⁶
Total							

6. **Periodical Checking of Materials:** To control the issue of materials this is very much necessary that bin cards, store control records and store ledgers are checked regularly and if any discrepancy is found, proper corrective actions should be taken.
7. **Physical Stock Checking of Materials:** Physical stock checking in stores should be done to prevent materials loss, material damage and theft. This checking can be done weekly, monthly etc. Physical stock checking means the verification of actual quantity in stores. This checking should be done surprisingly or at random basis. If any discrepancy is found and corrective actions should be taken to reduce or eliminate them the possible reasons may be wear and tear of materials, absorption of moisture, evaporation, waste, breakage, theft or wrong recordings. This is assumed to be the best method of inventory control.

Inventory Storing Procedure

Inventory storing procedure is an important part of inventory control management or materials management. Following procedure is followed in inventory storing:

1. **Receipt of Material in Store:** The storekeeper receives the material alongwith the goods received note from the receiving section. The material are then classified according to the nature of the material. The material should be arranged in bins especially meant for the materials. A bin card is attached with each bin or rack displaying the identification mark or code, minimum, maximum and ordering levels of materials and receipts, issues and balance of materials in hand, so that the exact position may be known at any time whenever desired.

Specimen of Bin Card

ABC Co. Ltd.

BIN CARD

Description	Maximum Level
Material code	Minimum Level
Location code	Danger Level
Bin No.	Ordering Level
Store Ledger Folio No.	Re-order Quantity

Receipts			Issues			Balance	Audit	
Date	G.R.N. No.	Qty.	Date	Rege.	Qty.	Qty.	Date	Initial

2. **Issue of Material from Store:** The store undertakes the responsibility of issuing the material to the using departments. In order to prevent malpractices, the materials must be issued only against the properly authorised requisition slips. These requisition must be properly checked and scrutinised to avoid overissue of materials. All requisition received must be posted immediately or daily on the bin cards and on the stock control cards. Generally three copies of requisition slips are prepared — first two copies are given to the stores and third copy kept with the demanding department. Store incharge keeps one copy of requisition slip for himself and other copy he sent to accounts department.

3. **Return of Material to Store:** If a department uses less material to its demand then it return the material to stores. Goods return slips are sent along with the materials. The same specifications and details of materials are given in goods return slips as they were mentioned in requisition slips. Three copies of goods return slips are prepared. First two copies are sent to stores department and third copy is kept by the goods returning department itself. Store keeper sent one copy to accounts department. The colour of both requisition slip and return slips are kept different to identify them easily.
4. **Transfer of Material:** The transfer of materials from one department to another department is generally not appreciated, because it creates problems in material control process. But sometime when there is emergency, the transfer of material from one department to other department is allowed. The department transferring the materials makes four copies of material transfer slips. First copy is sent to the needy department along with material. Second and third copies are sent to stores department and accounts department for their information.
5. **Material Abstract:** In big industries where the large quality of materials are received, issued and transferred daily, “material abstract” is prepared weekly or fortnightly to control the inventory. A physical verification of quantity in stores and other departments is done by material abstract.

If any discrepancy is found in physical verification of quantity in store or other department. It is brought into the notice of top management this type of check plays a very important role in inventory control. Thus material abstract is a summary of materials received, issued and transferred, for a given time period.

Check Your Progress 13.1

- 1 What are the main objectives of having Inventory Control?
2. Discuss ABC analysis techniques.

Notes: (a) Write your answer in the space given below.
(b) Please go through the lesson sub-head thoroughly you will get your answers in it.
(c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

13.7 MINIMUM STOCK LEVEL

The minimum stock level represents the lowest quantitative balance of materials in hand which must be maintained in hand at all times so that the assembly time may not be stopped on account of non-availability of materials.

The minimum stock level may be calculated by the use of the following formula:

$$\text{Minimum Stock Level} = \text{Re-ordered level} - (\text{Average rate of consumption} \times \text{Lead time})$$

These are as follows:

1. **Lead Time:** This is the time lag required to obtain the delivery of fresh supplies. If this time is more than the minimum inventory level will be high.
2. **Inland or Importable Inventory:** If the material is to be import then the lead time will be more implying minimum inventory level is to be kept high.
3. **Availability of Inventory:** If the material is not easily available then the minimum stock level to be kept high.
4. **Possibility of Interruption in Production:** If the production process is smooth then it is easy to determine the minimum stock level, but, if production is not smooth due to some reasons such as strike, power, etc. Then it is not easy to find out exact level of minimum stock.
5. **Nature of the Material:** Materials that are regularly stored must maintain a minimum level. If on customer's order a special item of material is to be purchased, no minimum level is required to be fixed for that.
6. **The Maximum Time Required from the Date of Order to the Date of Actual Delivery:** It is known as the Lead Time. The longer the lead time the lower is the minimum level, provided the reorder point remains constant.
7. **Rate of Consumption of the Material:** The minimum rate, the maximum rate and the normal rate of consumption are to be taken into consideration.

13.8 MAXIMUM STOCK LEVEL

Maximum stock level represents the maximum quantity of inventory which can be kept in store at any time. This quantity is fixed keeping in view of disadvantages of over-stocking.

Computation of Maximum Stock Level: The following formula is used:

$$\text{Maximum Level} = \text{Re-order level} + \text{Re-order quantity} - \text{Minimum consumption} \times \text{Minimum Re-order period}$$

Or

$$= \text{Re-order level} + \text{Re-order Quantity} - (\text{Average rate of usage} \times \text{Lead Time})$$

Factors Affecting the Maximum Stock Level

These are as follows:

1. Rate of consumption of the material.
2. The lead time.
3. The maximum requirement of the material at any point of time.
4. **Nature of the Material:** The materials which deteriorate quickly are stored as minimum as possible.
5. Storage space available for the material.
6. **Price Economy:** Seasonal materials are cheap during the harvesting reasons. So maximum purchase is made during that season and as a result the maximum level is high.
7. Cost of storage and insurance.
8. Cost of the material and the finance available. When the material is costly the maximum level is likely to be relatively low. If the price is likely to go up maximum level should be high.

9. **Inventory Turnover:** In case of slow moving materials the maximum level is low and in case of quick moving material it is high.
10. **Nature of Supply:** If the supply is uncertain the maximum level should be as high as possible.
11. **Economic Order Quantity (EOQ):** Maximum level largely depends in economic order quantity, because unless otherwise contra indicated the economic order quantity decides the quantity ordered and hence decides the maximum level.

13.9 ORDERING LEVEL OR RE-ORDER LEVEL

This is the fixed point between the maximum stock level and minimum stock levels at which time the order for next supply of materials from vendor is to be done.

This is mainly depends upon two factors:

1. Rate of Maximum usage.
2. Maximum Re-order period or Maximum Delivery Time.

Computation of Ordering Level or Re-order Level. The formula is as follows:

Ordering level or Re-order level

$$= \text{Maximum usage per day} \times \text{Maximum Re-order period or Maximum Delivery Time}$$

Or

$$= \text{Maximum Level} + (\text{Normal usage of Average rate of consumption} \times \text{Average Re-order period or Average Delivery Time})$$

Assumptions of Re-order point

1. The time of delivery remains fixed.
2. Load time remains fixed.
3. The average rate of consumption of materials does not changes.

13.10 AVERAGE STOCK LEVEL

Average stock level is the average quantity of stock for a given time of period.

Computation of Average Stock Level. The formula is as follows:

$$\text{Average stock level} = \frac{1}{2} [\text{Minimum Level} + \text{Maximum level}]$$

$$\text{or} \quad \text{Average stock level} = \text{Minimum level} + \frac{1}{2} [\text{Re-order Quantity}]$$

$$\text{or} \quad \text{Average stock level} = \frac{2(\text{Minimum level}) + \text{Re-order quantity}}{2}$$

13.11 DANGER LEVEL

In addition to the minimum, the maximum and recording levels there is another level called Danger Level. This level is below the minimum level and when the actual stock reaches this level immediate measure is to be taken to replenish stock. When the normal lead time is not available, the purchase quantity cannot be accurately determined. So, it is fixed in such a way that the actual stock does not fall below danger level by the actual lead time. This means, that the minimum level contains a cushion to cover contingencies.

Some concerns fix danger level below the re-ordering level but above the minimum level. If action for purchase is taken as soon as the stock reaches the re-ordering level, the danger level bears no importance except that, when the stock reaches the danger level (but not yet the minimum level) a reference may be made to the purchase department to ensure that delivery is received before the actual stock reaches the minimum level.

When the danger level is fixed below the minimum, it being reaches by the actual stock, the defect in the system is identified and corrective measure becomes necessary. When the danger level is fixed above the minimum, it being reached by the actual stock, preventive measure is to be taken so that the stock may not go below the minimum level.

It is the point or level of stock which the material stock should never be allowed to reduce. It is generally a level below the minimum level. As soon as the stock of material reaches this point, urgent action is needed for replenishment of stock.

Determination of Danger Level. This done as follows:

Danger Level = Two days of normal consumption

Re-order Quantity: The quantity which is ordered at re-order point is called re-order quantity. This is determined on the basis of minimum stock level and maximum stock level. This is normally used in notation of economic order quantity.

Check Your Progress 13.2

- 1 Differentiate:
 - (a) Minimum stock level and Maximum stock level
 - (b) Average Stock level and Danger Stock level
 2. What is Re-order level? Write assumptions for ascertaining Re-order point.
- Notes:**
- (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-
-

Numerical Solved Examples

Example 1: Calculate (i) Re-order Level; (ii) Minimum Level; and (iii) Maximum Level for each Component A and from the following information:

Normal Usage	50 Units per week each
Minimum Usage	25 Units per week each
Maximum Usage	75 Units per week each
Re-order Quantity	A: 300 Units; B : 500 Units
Re-order Period	A : 4 to 6 weeks; B : 2 to 4 weeks

Solution:

- (i) Re-order Level = Maximum Usage × Maximum Re-order Period
For Component A = $75 \times 6 = 450$ Units
For Component B = $75 \times 4 = 300$ Units
- (ii) Minimum Level = Re-order Level – (Normal Usage × Average Re-order Period)
For Component A = $450 - (50 \times 5) = 200$ Units
For Component B = $300 - (50 \times 3) = 150$ Units
- Note:** Average Re-order Period for Component A = $\frac{4+6}{2} = 5$
Average Re-order Period for Component B = $\frac{2+4}{2} = 3$
- (iii) Maximum Level = (Re-order Level + Re-order Quantity – (Minimum Usage × Minimum Re-order Period))
For Component A = $(450 + 300) - (25 \times 4)$
= 650 Units
For Component B = $(300 + 500) - (25 \times 2)$
= 750 Units

Example 2: From the following particulars, calculate: (a) Re-order Level (b) Minimum Level, (c) Maximum Level, (d) Average Level:

Normal Usage	100 units per day
Minimum Usage	60 units per day
Maximum Usage	130 units per day
Economic Order Quantity	5,000 units
Re-order Period	25 to 30 days

Solution:

- (a) Re-order Level = Maximum Usage × Maximum Re-order Period
= $130 \times 30 = 3,900$ units
- (b) Minimum Level = Re-order Level – (Normal Usage × Average Re-order Period)

= $3,900 - (100 \times 27.5) = 1,150$ units

Note: Average Re-order Period = $\frac{25+30}{2} = 27.5$ days
- (c) Maximum Level = (Re-order Level + Re-order Quantity Or EOQ)
– (Minimum Usage × Minimum Re-order Period)
= $(3,900 + 5,000) - 60 \times 25$
= 7,400 Units
- (d) Average Level = $\frac{\text{Minimum Level} + \text{Maximum Level}}{2} = 27.5$

= $\frac{1,150 + 7,400}{2} = 4,275$ Units

Example 3: A manufacturer buys costing equipment from out side suppliers Rs. 30 per unit. Total annual needs are 800 units. The following data is available:

Annual Return on Investment 10%

Rent, Insurance etc. per unit per year Re. 1

Cost of Placing an order Rs. 100

Determine Economic Order Quantity.

Solution:

$$EOQ = \sqrt{\frac{2 \times R \times C_p}{C_h}}$$

Where, EOQ = Economic Order Qunatity

R = Annual Requirement of Inventory

C_p = Cost of placing an order

C_h = Annual holding Or Carrying cost per unit per year.

Given : R = 800 units, C_p = Rs. 100, C_h = Rs. 4

$$EOQ = \sqrt{\frac{2 \times 800 \times 100}{4}} = \sqrt{40,000} = 200 \text{ Equipments}$$

$$= 10 \% \text{ of Rs. } 30 + \text{Re. } 1 = \text{Rs. } 3 + \text{Re. } 1 = \text{Rs. } 4.$$

Example 4: Fair Deal Limited uses Rs. 1,00,000 materials per year. The administration cost per purchase is Rs. 100 and the carrying cost is 20% of the average inventory. The company has a purchase policy on the basis of economic order quantity but has been offered a discount of 0.5% in the case of purchase five times per year. Advise the company whether it should accept new offer or not?

Solution: Given: R(in Rs.) = 1,00,000, C_p = Rs. 100, P = Re. 1.00,

$$C_h = 1.00 \times 20\% = \text{Re. } 0.20$$

$$\text{E.O.Q. (in Rs.)} = \sqrt{\frac{2 \times R \times C_p}{C_h}}$$

$$= \sqrt{\frac{2 \times 1,00,000 \times 100}{0.20}}$$

$$= \sqrt{10,00,00,000}$$

$$= \text{Rs. } 10,000$$

Total Inventory Cost in case of each order is placed of Rs. 10,000:"

(i) Cost of Materials Rs. 1,00,000

(ii) Ordering Cost = $\frac{R}{q_0} \times C_p = \frac{1,00,000}{10,000} \times 100$ Rs. 1,000

(iii) Carrying Cost = $\frac{q_0}{2} \times C_h = \frac{10,000}{2} \times 0.2$ Rs. 1,000

Total Cost	<u>Rs. 1,02,000</u>
------------	---------------------

Total Cost in case of each order is placed or Rs. 19,900 i.e., Rs. 20,000 – 0.5% discount:

(i) Cost of Materials	$(19,900 \times 5)$	Rs. 99,500.00
(ii) Ordering Cost	$= \left(\frac{R}{q_0} \times C_p \right)$	
	$= \frac{99,500}{19,900} \times 100$	500.00
(iii) Carrying Cost	$= \left(\frac{q_0}{2} \times C_h \right)$	
	$= \frac{19,900}{2} \times 0.199$	1,980.05

Total Inventory Cost		1,01,980.05

[**Note:** Here P = Re. 1, 0.5% or Re. 1 = Re. 0.95, $C_h = 0.95 \times 20\% = \text{Re. } 0.199$]

On the basis of above analysis the offer should be accepted as it will save Rs. 1,02,000 – 1,01,980.05 = Rs. 19.95.

Example 5: A pharmaceutical factory consumes annually 6,000 kgms. of a chemical costing Rs. 5 per kgm. Placing each order costs Rs. 25 and the carrying cost is 6% per year per kgm. of average inventory. Find the Economic Order Quantity and the total inventory cost.

The factory works for days in a year. If the procurement time is 15 days and safety stock 200 kgms., find the re-order point and maximum and average inventories levels.

If the supplier offers a discount of 5% on the cost price for a single order of annual requirement, should the factory accept it?

Solution: Given: R = 6,000 kgms.; P = Rs. 5 per kgm. $C_p = \text{Rs. } 25$; $CH = 6\%$ per kgm. per year of average inventory; No. of working days in a year = 300; Procurement time = 15 days; Safety Stock = 200 kgms.

$$\begin{aligned} E.O.Q. &= \sqrt{\frac{2 \times R \times C_p}{C_h}} \\ &= \sqrt{\frac{2 \times 60,000 \times 25}{.30}} = \sqrt{\frac{3,00,000}{.30}} \\ &= \sqrt{10,00,000} = 1,000 \text{ kgms.} \end{aligned}$$

$$C_h = 6\% \text{ of Average inventory i.e., } \frac{5 \times 6}{100} = \text{Re. } 30$$

$$\begin{aligned} T.I.C &= (R \times P) + \left(\frac{R}{q_0} \times C_p \right) + \left(\frac{q_0}{2} \times C_h \right) \\ &= (6,000 \times 5) + \left(\frac{6,000}{1,000} \times 25 \right) + \left(\frac{1,000}{2} \times .30 \right) \\ &= 30,000 + 150 + 140 = \text{Rs. } 30,000 \end{aligned}$$

$$\text{Re-order Point} = \left(\frac{R}{\text{No. of Working days}} \times \text{Procurement time} \right) + \text{Safety Stock}$$

$$= \left(\frac{6,000}{300} \times \frac{15}{1} \right) + 200$$

$$= 300 + 200 = 500 \text{ kgms.}$$

$$\begin{aligned}\text{Maximum Stock Level} &= (\text{Re-order Point} + \text{Re-order Quantity or EOQ}) \\ &\quad - (\text{Minimum Usage} \times \text{Minimum Re-order Period}) \\ &= (500 + 1,000) - (20 \times 15) \\ &= 1,500 - 300 = 1,200 \text{ kgms.}\end{aligned}$$

or Maximum Stock Level = $q_0 + \text{Safety Stock}$
 $= 1,000 + 200 = 1,200 \text{ kgms.}$

$$\begin{aligned}\text{Minimum Stock Level} &= \text{Re-order Level} - (\text{Normal usage} \times \text{Average Re-order period}) \\ &= 500 - (20^* \times 15) \\ &= 500 - 300 = 200 \text{ kgms.}\end{aligned}$$

$$\begin{aligned}{}^* \text{Normal Usage} &= \frac{R}{\text{No. of Working days}} \\ &= \frac{6,000}{300} = 20 \text{ kgms.}\end{aligned}$$

$$\begin{aligned}\text{Average Stock Level} &= \left(\frac{\text{Minimum Stock Level} + \text{Maximum Stock Level}}{2} \right) \\ &= \frac{200 + 1,200}{2} = \frac{1,400}{2} = 700 \text{ kgms.}\end{aligned}$$

Or Average Stock Level = $\frac{q_0}{2} + \text{Safety Stock}$
 $= \frac{1,000}{2} + 200 = 700 \text{ kgms.}$

TIC if a single order of 6,000 kgms is placed:

Given: P = Rs. 5 – 5% of Rs. 5 i.e., $5 - .25 = \text{Rs. } 4.75$

$$C_H = 6\% \text{ of Average Inventory i.e., } 4.75 \times \frac{6}{100} = \text{Rs. } 285;$$

$$C_p = \text{Rs. } 25; R = 6,000 \text{ kgms}; q_0 = 6,000 \text{ kgms.}$$

$$\begin{aligned}\text{TIC} &= (R \times P) + \left(\frac{R}{q_0} \times C_p \right) + \left(\frac{q_0}{2} \times C_H \right) \\ &= (6,000 \times 4.75) + \left(\frac{6,000}{6,000} \times 25 \right) + \left(\frac{6,000}{2} \times .25 \right) \\ &= 28,500 + 25 + 855 = \text{Rs. } 29,380.\end{aligned}$$

The company should accept the offer of 5% discount in purchase price by placing a single order of 6,000 kgms. because the total inventory cost in this case is less by Rs. $30,300 - \text{Rs. } 29,380 = \text{Rs. } 920$ as compared to total inventory cost without discount offer.

Example 6: A trading company expects to sell 15,000 mixers during the coming year. The cost per mixer is Rs. 200. The cost of storing a mixer for 1 year is Rs. 5 and the ordering cost is Rs. 540 per order. Find the Economic Order Quantity. Would it be profitable to the company to accept a discount offer of 30% on a single order per year. The storing cost continuing to be Rs. 5 per mixer per year.

Solution:

$$E.O.Q = \sqrt{\frac{2 \times R \times C_p}{C_h}}$$

Given $R = 15,000$ units, $C_p = 540$, $C_h = \text{Rs. } 5$

$$E.O.Q = \sqrt{\frac{2 \times 15,000 \times 540}{5}}$$

$$= \sqrt{32,40,000} = 1,800 \text{ units}$$

Total Inventory Cost if $q_0 = 1,800$ units:

$$\begin{aligned} T.I.C &= (R \times P) + \left(\frac{R}{q_0} \times C_p \right) + \left(\frac{q_0}{2} \times C_h \right) \\ &= (15,000 \times 200) + \left(\frac{15,000}{1,800} \times 540 \right) + \left(\frac{1,800}{2} \times 5 \right) \\ &= 30,00,000 + 4,500 + 4,500 \\ &= \text{Rs. } 30,09,000 \end{aligned}$$

T.I.C. if a single order is placed at 30% discount in price:

$$\begin{aligned} T.I.C &= (R \times P) + \left(\frac{R}{q_0} \times C_p \right) + \left(\frac{q_0}{2} \times C_h \right) \\ &= (15,000 \times 140) + \left(\frac{15,000}{15,000} \times 540 \right) + \left(\frac{15,000}{2} \times 5 \right) \\ &= 21,00,000 + 540 + 37,500 \\ &= \text{Rs. } 21,38,040 \end{aligned}$$

The company should accept the offer of 30% discount as it will save $\text{Rs. } 30,09,000 - 21,38,040 = \text{Rs. } 8,70,960$.

Example 7: A manufacturer requires 1,000 units of a raw material, per month. The ordering cost is Rs. 15 per order. The carrying cost in addition to Rs. 2 per unit, is estimated to be 15% of average inventory per unit per year. The purchase price of the raw material is Rs. 10 per unit. Find the Economic Lot Size and the total cost.

The manufacturer is offered as 5% discount in purchase price for order for 2,000 units or more but less than 5,000 units. A further 2% discount is available for order of 5,000 or more units. Which of the three ways of purchase he should adopt?

Solution:

Given: $R = 1,000$ units per month or $12,000$ units per annum;

$C_p = \text{Rs. } 15$ per order;

- P = (i) Rs. 10 per unit in case of order for less than 2,000 units.
(ii) Rs. 10 – 5% of Rs. 10 i.e., Rs. 9.50 in case of order for 2,000 or more units but less than 5,000 units.
(iii) Rs. 10 – 7% of Rs. 10 i.e., Rs. 9.30 in case of order for 5,000 or more units.

- C_H = (i) Rs. $2 + 15\%$ of Rs. 2 of Average inventory i.e., Rs. $2 + 1.50 =$ Rs. 3.50 per unit per annum in case of order for less than 2,000 units.
(ii) Rs. $2 + 15\%$ of Rs. 9.50 = Rs. $2 + 1.425 =$ Rs. 3.425 per unit per annum in case of order for 2,000 units or more but less than 5,000 units.
(iii) Rs. $2 + 15\%$ of Rs. 9.70 = Rs. $2 + 1.395 =$ Rs. 3.395 per unit per annum in case of order for 5,000 or more units.

Alternative I: In case of order for less than 2,000 units:

$$\begin{aligned} \text{E.O.Q. } (q_0) &= \sqrt{\frac{2 \times R \times C_p}{C_H}} \\ &= \sqrt{\frac{2 \times 12,000 \times 15}{3.50}} = \sqrt{\frac{3,60,000}{3.50}} \\ &= \sqrt{1,02,857} = 320.7 \text{ or } 321 \text{ units} \end{aligned}$$

$$\begin{aligned} \text{T.I.C.} &= (R \times P) + \left(\frac{R}{q_0} \times C_p \right) + \left(\frac{q_0}{2} \times C_H \right) \\ &= (12,000 \times 10) + \left(\frac{12,000}{321} \times 15 \right) + \left(\frac{321}{2} \times 3.50 \right) \\ &= 1,20,000 + 561 + 562 \\ &= \text{Rs. } 1,21,123 \text{ (nearest to Rupee)} \end{aligned}$$

Alternative II: In case of order for 2,000 or more units but less than 5,000 units:

$$\begin{aligned} \text{E.O.Q. } (q_0) &= \sqrt{\frac{2 \times R \times C_p}{C_H}} \\ &= \sqrt{\frac{2 \times 12,000 \times 15}{3.425}} = \sqrt{\frac{3,60,000}{3.425}} \\ &= \sqrt{1,05,109.45} = 324 \text{ units} \end{aligned}$$

As the Economic Lot size (324 units) is less than minimum ordering quantity (2,000 units), the company should order at least 2,000 units to get 5% discount in purchase price

Thus, T.I.C. if $q_0 = 2000$ units:

$$\begin{aligned} \text{T.I.C.} &= (R \times P) + \left(\frac{R}{q_0} \times C_p \right) + \left(\frac{q_0}{2} \times C_H \right) \\ &= (12,000 \times 9.50) + \left(\frac{12,000}{2,000} \times 15 \right) + \left(\frac{2,000}{2} \times 3.425 \right) \\ &= 1,14,000 + 90 + 3,425 = \text{Rs. } 1,17,515. \end{aligned}$$

Alternative III: In case of orders of 5,000 or more units:

$$\text{Economic Lot size } (q_0) = \sqrt{\frac{2 \times R \times C_p}{C_h}}$$

$$= \sqrt{\frac{2 \times 12,000 \times 15}{3.395}} = \sqrt{\frac{3,60,000}{3.395}}$$

$$= \sqrt{1,06,038.29} = 325.6$$

Or $= 326 \text{ units}$

As the Economic Lot Size (326 units) is less than the minimum ordering quantity 5,000 units, the company should order at least 5,000 units to get 7% discount in purchase price.

Thus T.I.C. if $q_0 = 5,000$ units:

$$\begin{aligned} \text{T.I.C.} &= (R \times P) + \left(\frac{R}{q_0} \times C_p \right) + \left(\frac{q_0}{2} \times C_h \right) \\ &= (12,000 \times 9.30) + \left(\frac{12,000}{5,000} \times 15 \right) + \left(\frac{5,000}{2} \times 3.395 \right) \\ &= 1,11,600 + 36 + 8,487.50 = 1,20,123.50 \end{aligned}$$

On the basis of above analysis we find that the T.I.C. is minimum (Rs. 1,17,515) in second alternative. Hence the company should adopt this alternative.

13.12 LET US SUM UP

In this chapter the more emphasis is given on the role of inventory control. The certain technique such as ABC Analysis have been illustrated with the help of specimen format and illustration.

13.13 LESSON-END ACTIVITIES

1. Successful well-organised business relies heavily on the inventory control to make certain they have adequate inventory levels to satisfy their customers. As we know that inventory control offers comprehensive reporting capabilities to keep you on top of inventory status. Take any manufacturing company like Pepsi or Coca-Cola which can help to bring about the creation of new or improved purchasing policies, sales policies, pricing methods and even enhanced customer service.
2. Take a case of a automobile industry where the inventory control had played a vital role like quickly locate parts, product lines, purchase orders, account payable vendor and general ledger account.

13.14 KEYWORDS

- Lead Time** : Time between ordering & receiving the good.
- Inventory Control System** : Is a technique to maintain inventory at a desired level.
- Maximum Level** : Level of inventory beyond which inventory is not allowed.
- Minimum Level** : Level of inventory beyond which inventory is allowed.

Opportunity Cost	:	The next best alternative cost.
Reorder Level	:	The stock level which is sufficient for the lead time consumption.

Inventory Model

13.15 QUESTIONS FOR DISCUSSION

1. **Write True or False against each statement:**
 - (a) Inventory can be defined as the state of goods.
 - (b) ABC analysis ensure the close control over the items of A, B and C categories.
 - (c) Head line represent the maximum quantity of inventory.
 - (d) Material control help to minimise loose by obsolescence.
2. **Write short notes on following:**
 - (a) Re-order level
 - (b) Minimum stock level
 - (c) Maximum stock level
 - (d) Danger level
 - (e) Average stock level
 - (f) EOQ
3. **Briefly comment on the following statements:**
 - (a) Modern technique for controlling the inventory in a value item analysis known as ABC.
 - (b) When the material is costly the maximum level is likely to be relatively low?
 - (c) Danger level is fixed above the minimum.
 - (d) There should be well organised issuing system of material.

13.16 TERMINAL QUESTIONS

1. What do you understand by inventory control?
2. Discuss the objectives of inventory control.
3. Discuss the various factors which determine the level of inventory control.
4. What is an ABC analysis? What are the steps in ABC analysis?
5. Explain the process of inventory control with example.
6. In manufacturing a commodity two components X and Y are used as follows:

Normal usage	100 units per week each
Minimum usage	50 units per week each
Maximum usage	150 units per week each
Ordering Quantities	X : 600 units; Y : 1,000 units
Delivery Period	X : 4 to 6 weeks Y : 2 to 4 weeks
7. From the following information determine the Re-order point, Minimum Stock Level and Maximum Stock Level:
 - (a) Minimum consumption 500 units per day
 - (b) Maximum consumption 875 units per day

- (c) Normal consumption 625 units per day
 - (d) Re-order Quantity 8,800 units
 - (e) Minimum period for receiving goods 7 days
 - (f) Maximum period for receiving goods 15 days
 - (g) Normal period for receiving goods 10 days
8. A manufacturer's requirement for raw materials is 12,800 kgms. per annum. The purchase price of it is Rs. 50 per kgm. Ordering cost is Rs. 100 per order and carrying cost is 8% of average inventory. The manufacturer can procure its annual requirement of raw material higher in one single lot or by ordering of 400, 800, 1600 or 3,200 kgms. quantity. Find which of these order quantities is the Economic Order Quantity using tabular method.
9. The annual requirement of a product in a firm is 1,000 units. The purchase price per unit is Rs. 50; ordering cost is Rs. 150 per order and the carrying cost per unit of average of inventory is 15%. The firm can procure its annual requirement either in one single lot or in various alternative lots of 100, 200, 250 or 500 units. Determine the Economic Order Quantity by Graphical method and with the help of the three curves, show at EOQ level ordering and carrying costs are equal and total cost is minimum.
10. Calculate Economic Order Quantity from the following information by using Tabular method, Graphical method and mathematical method:
- | | |
|----------------------------|--------------------------|
| Annual usage | 10,000 units |
| Buying cost per order | Rs. 10 |
| Cost per unit | Rs. 50 |
| Cost of carrying inventory | 10% of Average Inventory |
11. A company requires annually 12,000 lbs. of a chemical which costs Rs. 250 per lb. Placing each order costs the company Rs. 22.50, and the carrying cost is 15% of the cost of average inventory per annum.
- (i) Find Economic Order Quantity and total expenses on the chemical.
 - (ii) If in addition, the company decides to maintain a stock of 300 lbs. find the maximum as well as average inventory.
12. Calculate the Economic Order Quantity from the following information. Also state what will be the number of orders during the whole year:
- | | |
|-----------------------------------|--------------------------|
| Requirement of material per annum | 1,250 units |
| Cost of material per unit | Rs. 200 |
| Cost of placing per order | Rs. 100 |
| Holding cost per unit per annum | 8% of average inventory. |
13. A manufacturer's requirement for a raw material is 2,000 units per year. The ordering costs are Rs. 10 per order while carrying costs are 16 paise per year per unit of a average inventory. The purchase price of raw material is Re. 1 per unit.
- (a) Find the Economic Order quantity and the total inventory cost.
 - (b) If a discount of 5% is available for orders of 1,000 units, should the manufacturer accepts this offer?
- (The carrying cost per unit per annum remains unchanged.)
14. A business unit expect to sell 60,500 units of a commodity during the coming year. The ordering cost per order is Rs. 840 and the cost per unit of the commodity is Rs.

200. The carrying cost per unit per annum is 0.5% of the average inventory. Find out Economic Order Quantity. Would it be profitable to the business unit to accept a discount offer of 1% on a single order per year. In this case the storing cost per unit per year will increase to 0.75% of the average inventory.
15. A manufacturer requires 2,500 units of a raw material per month. The ordering cost is Rs. 20 per order. The carrying cost in addition to Rs. 3 per unit, is estimated to be 10% of average inventory per unit per year. The purchase price of the raw material is Rs. 4 per unit. Find the Economic Lot Size and the Total Inventory Cost. The manufacturer is offered a discount in purchase price for order of 1,000 units or more but less than 2,000 units. A further discount is available for orders of 2,000 or more units. Which of the three ways of purchase he should adopt?

13.17 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

3. (a) True (b) True (c) False (d) True

13.18 SUGGESTED READINGS

M.P. Gupta & J.K. Sharma, *Operation Research for Management*, National Publishing House.

Mustafi C.K., *Operation Research Methods & Practice*, Wiley Eastern Ltd.

Peterson & E.A. Silver, *Decision System for Inventory Management Production Planning*, Wiley New York.

Stan M.K.D.W. Miller, *Inventory Control Theory & Practices*, Prentice Hall of India.

Tana H.A., *Operation Research Introduction*, Macmillan Publishing Company.

LESSON

14

GAME THEORY

CONTENTS

- 14.0 Aims and Objectives
 - 14.1 Introduction
 - 14.2 Two-person Zero-sum Game
 - 14.3 Pure Strategies: Game with Saddle Point
 - 14.4 Mixed Strategies: Games without Saddle Point
 - 14.5 Dominance Property
 - 14.6 Solving Problem on the Computer with TORA
 - 14.7 Solving LP Model Games Graphically using Computer
 - 14.8 Let us Sum Up
 - 14.9 Lesson-end Activity
 - 14.10 Keywords
 - 14.11 Questions for Discussion
 - 14.12 Terminal Questions
 - 14.13 Model Answers to Questions for Discussion
 - 14.14 Suggested Readings
-

14.0 AIMS AND OBJECTIVES

In this lesson we are going to discuss about the various strategies of players. This is done through games theory. A game refers to the situation in which two three players are competing.

14.1 INTRODUCTION

Game theory applies to those competitive situations which are technically known as “competitive games” or in general known as games. As the game is a competition involving two or more decisions makers each of whom is keen to win. The basic aim of this chapter is to study about how the optimal strategies are formulated in the conflict. Thus we can say that game theory is not related with finding an optimum or winning strategy for a particular conflict situation. Afterwards we can say that the theory of game is simply the logic of rational decisions. After reading this unit, you should be able to know how to take decision under the cut-throat competition and know that outcome of our business enterprise depends on what the competitor will do.

In today’s business world, decisions about many practical problems are made in a competitive situation, where two or more opponents are involved under the conditions of

competition and conflict situations. The outcome does not depend on the decision alone but also the interaction between the decision-maker and the competitor.

The objective, in theory, of games is to determine the rules of rational behaviour in game situations, in which the outcomes are dependent on the actions of the interdependent players. A game refers to a situation in which two or more players are competing. A player may be an individual, a group or an organization. Game Theory has formulated mathematical models that can be useful in decision-making in competitive situations. To get a better insight of the concept, we consider an example of a simple game.

Let us assume that there are only two car manufacturers, company A and company B. The two companies have market shares for their product. Company A is planning to increase their market share for the next financial year. The vice-president of company A has come up with two strategies. One strategy is to modify the outer shape of the car and to advertise on TV. Company B, knowing that if these strategies are adopted by company A, it may lead to decrease in its market share, develops similar strategies to modify the shape of their car and to advertise on TV. Table 14.1 below, gives the pay off if both the companies adopt these strategies.

Table 14.1: The Pay Off if Both Companies Modify Shape & Advertise on TV

		Company B	
		Modify shape	Advertise
Company A		Modify shape	4
		Advertise	-5

The pay off given is with respect to company A and represents company A. Company B's pay off is the opposite of each element. For example, it means that for modification strategy, Company A wins 4 and company B loses 4.

In a game, each player has a set of strategies available. A strategy of a player is the list of all possible actions (course of action) that are taken for every pay-off (outcome). The players also know the outcome in advance. The players in the game strive for optimal strategies. An optimal strategy is the one, which provides the best situation (maximum pay-off) to the players.

Payoff Matrix: Company A has strategies A_1, A_2, \dots, A_m , and Company B has strategies B_1, B_2, \dots, B_n . The number of pay-offs or outcomes is $m \times n$. The pay-off a_{mn} represents company A's gains from Company B, if company A selects strategy m and company B selects strategy n. At the same time, it is a loss for company B ($-a_{mn}$). The pay-off matrix is given (Table 14.2) with respect to company A.

The game is zero-sum because the gain of one player is equal to the loss of other and vice-versa.

Table 14.2: Pay-off Matrix

Company B Strategies

		B ₁	B ₂	B ₃	B _n	
Company A		A ₁	a ₁₁	a ₁₂	a ₁₃	a _{1n}
		A ₂	a ₂₁	a ₂₂	a ₂₃	a _{2n}
Strategies	A ₃	a ₃₁	a ₃₂	a ₃₃	a _{3n}	
	
	
A _m	a _{m1}	a _{m2}	a _{m3}	a _{mn}		

14.2 TWO-PERSON ZERO-SUM GAME

In a game with two players, if the gain of one player is equal to the loss of another player, then the game is a two person zero-sum game.

A game in a competitive situation possesses the following properties:

- i. The number of players is finite.
- ii. Each player has finite list of courses of action or strategy.
- iii. A game is played when each player chooses a course of action (strategy) out of the available strategies. No player is aware of his opponent's choice until he decides his own.
- iv. The outcome of the play depends on every combination of courses of action. Each outcome determines the gain or loss of each player.

14.3 PURE STRATEGIES: GAME WITH SADDLE POINT

The aim of the game is to determine how the players must select their respective strategies such that the pay-off is optimized. This decision-making is referred to as the **minimax-maximin principle** to obtain the best possible selection of a strategy for the players.

In a pay-off matrix, the minimum value in each row represents the minimum gain for player A. Player A will select the strategy that gives him the maximum gain among the row minimum values. The selection of strategy by player A is based on maximin principle. Similarly, the same pay-off is a loss for player B. The maximum value in each column represents the maximum loss for Player B. Player B will select the strategy that gives him the minimum loss among the column maximum values. The selection of strategy by player B is based on minimax principle. If the maximin value is equal to minimax value, the game has a saddle point (i.e., equilibrium point). Thus the strategy selected by player A and player B are optimal.

Example 1: Consider the example to solve the game whose pay-off matrix is given in Table 14.3 as follows:

Table 14.3: Game Problem

		Player B	
		1	2
		1	3
Player A	1	1	
	2	-1	6

The game is worked out using minimax procedure. Find the smallest value in each row and select the largest value of these values. Next, find the largest value in each column and select the smallest of these numbers. The procedure is shown in Table 14.4.

Table 14.4: Minimax Procedure

		Player B		Row Min
		1	2	
		1	3	1
Player A →	1	1		1
	2	-1	6	
		1	6	
Col Max →		1	6	

If Maximum value in row is equal to the minimum value in column, then saddle point exists.

Game Theory

$$\text{Max Min} = \text{Min Max}$$

$$1 = 1$$

Therefore, there is a saddle point.

The strategies are,

Player A plays Strategy A_1 , (A_1 A_1).

Player B plays Strategy B_1 , (B_1 B_1).

Value of game = 1.

Example 2: Solve the game with the pay-off matrix for player A as given in Table 14.5.

Table 14.5: Game Problem

		Player B		
		B_1	B_2	B_3
Player A	A_1	-4	0	4
	A_2	1	4	2
	A_3	-1	5	-3

Solution: Find the smallest element in rows and largest elements in columns as shown in Table 14.6.

Table 14.6: Minimax Procedure

		Player B			
		B_1	B_2	B_3	Row min
Player A	A_1	-4	0	4	-4
	A_2	1	4	2	(1)
	A_3	-1	5	-3	-3
$\xrightarrow{\text{Column Max}}$		(1)	5	4	

Select the largest element in row and smallest element in column. Check for the minimax criterion,

$$\text{Max Min} = \text{Min Max}$$

$$1 = 1$$

Therefore, there is a saddle point and it is a pure strategy.

Optimum Strategy:

Player A A_2 Strategy

Player B B_1 Strategy

The value of the game is 1.

Example 3: Check whether the following game is given in Table 14.7, determinable and fair.

Table 14.7: Game Problem

		Player B	
		1	2
Player A		1	7
		2	0
			8

Solution: The game is solved using maximin criteria as shown in Table 14.8.

Table 14.8: Maximin Procedure

		Player B		Row Min
		1	2	
Player A		1	7	0
		2	0	8
		Column Max		8

Max Min \neq Min Max

i.e., $0 \neq 7$

The game is strictly neither determinable nor fair.

Example 4: Identify the optimal strategies for player A and player B for the game, given below in Table 14.9. Also find if the game is strictly determinable and fair.

Table 14.9: Game Problem

		Player B		Row Min
		1	2	
Player A		1	4	0
		2	1	-3
		Col Max		-3
		4		0

Max Min = Min Max

$0 = 0$

The game is strictly determinable and fair. The saddle point exists and the game has a pure strategy. The optimal strategies are given in Table 14.10 (a, b).

Table 14.10: Optimal Strategies

$$(a) S_A \begin{pmatrix} 1 & 2 \\ p_1 & p_2 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad (b) S_B \begin{pmatrix} 1 & 2 \\ q_1 & q_2 \\ 0 & 1 \end{pmatrix}$$

Example 5 : Solve the game with the pay off matrix given in Table 14.11 and determine the best strategies for the companies A and B and find the value of the game for them.

Table 14.11: Game Problem

Game Theory

		Company B		
		2	4	2
Company A		1	-5	-4
		2	6	-2

Solution: The matrix is solved using maximin criteria, as shown in Table 14.12 below.

Table 14.12: Maximin Procedure

		Company B			
		1	2	3	Row Min
Company A		1	2	2	②
		2	1	-5	-5
Column Max		②	6	②	

$$\text{Max Min} = \text{Min Max}$$

$$2 = 2$$

Therefore, there is a saddle point.

Optimum strategy for company A is A_1 and

Optimum strategy for company B is B_1 or B_3 .

Check Your Progress 14.1

1. Discuss two-person zero-sum game.
2. What is minimax-minimin principle?

Notes: (a) Write your answer in the space given below.

- (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-
-

14.4 MIXED STRATEGIES: GAMES WITHOUT SADDLE POINT

For any given pay off matrix without saddle point the optimum mixed strategies are shown in Table 14.13.

Table 14.13: Mixed Strategies

		Player B	
		B ₁	B ₂
Player A		A ₁	a_{11} a_{12}
		A ₂	a_{21} a_{22}

Let p₁ and p₂ be the probability for Player A.

Let q₁ and q₂ be the probability for Player B.

Let the optimal strategy be S_A for player A and S_B for player B.

Then the optimal strategies are given in Tables 14.14 a & b.

Table 14.14 (a), (b): Optimum Strategies

$$(a) \quad S_A = \begin{pmatrix} A_1 & A_2 \\ p_1 & p_2 \end{pmatrix} \quad \text{and} \quad (b) \quad S_B = \begin{pmatrix} B_1 & B_2 \\ q_1 & q_2 \end{pmatrix}$$

p₁ and p₂ are determined by using the formulae,

$$p_1 = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})} \quad \text{and} \quad p_2 = 1 - p_1$$

$$q_1 = \frac{a_{22} - a_{12}}{(a_{11} + a_{22}) - (a_{12} + a_{21})} \quad \text{and} \quad q_2 = 1 - q_1$$

and the value of the game w.r.t. player A is given by,

$$a_{11}a_{22} - a_{12}a_{21}$$

Value of the game, v =

$$(a_{11} + a_{22}) - (a_{12} + a_{21})$$

Example 6: Solve the pay-off given Table 14.15 matrix and determine the optimal strategies and the value of game.

Table 14.15: Game Problem

		Player B	
		1	2
Player A		1	5 2
		2	3 4

Solution: Let the optimal strategies of S_A and S_B be as shown in Tables 14.16 (a, b).

Table 14.16(a) and (b): Optimal Strategies

$$(a) \quad S_A = \begin{pmatrix} A_1 & A_2 \\ p_1 & p_2 \end{pmatrix} \quad \text{and} \quad (b) \quad S_B = \begin{pmatrix} B_1 & B_2 \\ q_1 & q_2 \end{pmatrix}$$

The given pay-off matrix is shown below in Table 14.17.

Table 14.17: Pay-off Matrix or Maximin Procedure

Game Theory

		Player B		Row Min	
		1	2		
Player A	1	5	2	(2)	
	2	3	4	3	
Column Max		5	(4)	Max Min ≠ Min Max	
				3 ≠ 4	

Therefore, there is no saddle point and hence it has a mixed strategy.

Applying the probability formula,

$$p_1 = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$$

$$= \frac{4 - 3}{(5 + 4) - (2 + 3)} = \frac{1}{9 - 5} = \frac{1}{4}$$

$$p_2 = 1 - p_1 = 1 - \frac{1}{4} = \frac{3}{4}$$

$$q_1 = \frac{a_{22} - a_{12}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$$

$$= \frac{4 - 2}{(5 + 4) - (2 + 3)} = \frac{2}{9 - 5} = \frac{2}{4} = \frac{1}{2}$$

$$q_2 = 1 - q_1 = 1 - \frac{1}{2} = \frac{1}{2}$$

$$\text{Value of the game, } v = \frac{a_{11}a_{22} - a_{12}a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$$

$$= \frac{(5 \times 4) - (2 \times 3)}{(5 + 4) - (2 + 3)}$$

The optimum mixed strategies are shown in Table 14.18 (a, b) below.

Table 14.18(a) and (b): Optimum Mixed Strategies

$$(a) S_A = \begin{pmatrix} A_1 & A_2 \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} \quad \text{and} \quad (b) S_B = \begin{pmatrix} B_1 & B_2 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Value of the game = 3.5

14.5 DOMINANCE PROPERTY

In case of pay-off matrices larger than 2×2 size, the dominance property can be used to reduce the size of the pay-off matrix by eliminating the strategies that would never be selected. Such a property is called a dominance property.

Example 7: Solve the game given below in Table 14.19 after reducing it to 2×2 game:

Table 14.19: Game Problem

		Player B		
		1	2	3
Player A	1	1	7	2
	2	6	2	7
	3	5	1	6

Solution: Reduce the matrix by using the dominance property. In the given matrix for player A, all the elements in Row 3 are less than the adjacent elements of Row 2. Strategy 3 will not be selected by player A, because it gives less profit for player A. Row 3 is dominated by Row 2. Hence delete Row 3, as shown in Table 14.20.

Table 14.20: Reduced the Matrix by Using Dominance Property

		Player B		
		1	2	3
Player A	1	1	7	2
	2	6	2	7

For Player B, Column 3 is dominated by column 1 (Here the dominance is opposite because Player B selects the minimum loss). Hence delete Column 3. We get the reduced 2×2 matrix as shown below in Table 14.21.

Table 14.21: Reduced 2×2 Matrix

		Player B	
		1	2
Player A	1	1	7
	2	6	2

Now, solve the 2×2 matrix, using the maximin criteria as shown below in Table 14.22.

Table 14.22: Maximin Procedure

		Player B		Row Min	
		1	2		
Player A	1	1	7	1	
	2	6	2	(2)	
Column Max		(6)	7		

Max Min \neq Min Max

$2 \neq 6$

Therefore, there is no saddle point and the game has a mixed strategy.

Applying the probability formula,

$$p_1 = \frac{2-6}{(1+2)-(7+6)}$$

$$= \frac{-4}{3-13} = \frac{4}{10} = \frac{2}{5}$$

$$q_1 = 1 - \frac{2}{5} = \frac{3}{5}$$

$$q_1 = \frac{2-7}{(1+2)-(7+6)} = \frac{-5}{3-13} = \frac{5}{10} = \frac{1}{2}$$

$$q_2 = 1 - q_1 = 1 - \frac{1}{2} = \frac{1}{2}$$

$$\text{Value of the game, } v = \frac{(1 \times 2) - (7 \times 6)}{(1+2)-(7+6)} = \frac{2-42}{3-13} = \frac{40}{10} = 4$$

The optimum strategies are shown in Table 14.23 (a, b)

Table 14.23 (a, b): Optimum Strategies

$$(a) S_A = \begin{pmatrix} A_1 & A_2 & A_3 \\ \frac{2}{5} & \frac{3}{5} & 0 \end{pmatrix} \text{ and (b) } S_B = \begin{pmatrix} B_1 & B_2 & B_3 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

$$\text{Value of the game, } v = 4$$

Example 8: Is the following two-person zero-sum game stable? Solve the game given below in Table 14.24.

Table 14.24: Two-person Zero-sum Game Problem

		Player B			
		1	2	3	4
Player A	1	5	-10	9	0
	2	6	7	8	1
	3	8	7	15	1
	4	3	4	-1	4

Solution: Solve the given matrix using the maximin criteria as shown in Table 14.25.

Table 14.25: Maximin Procedure

		Player B				Row Min
		1	2	3	4	
Player A	1	5	-10	9	0	(-10)
	2	6	7	8	1	(1)
	3	8	7	15	1	1
	4	3	4	-1	4	-1
Column Max		8	7	15	4	(4)

Max Min \neq Min Max

$$3 \neq 1$$

Therefore, there is no saddle point and hence it has a mixed strategy.

The pay-off matrix is reduced to 2×2 size using dominance property. Compare the rows to find the row which dominates other row. Here for Player A, Row 1 is dominated by Row 3 (or row 1 gives the minimum profit for Player A), hence delete Row 1. The matrix is reduced as shown in Table 14.26.

Table 14.26: Use Dominance Property to Reduce Matrix (Deleted Row 1)

		Player B			
		1	2	3	4
Player A	2	6	7	8	1
	3	8	7	15	1
	4	3	4	-1	4

When comparing column wise, column 2 is dominated by column 4. For Player B, the minimum profit column is column 2, hence delete column 2. The matrix is further reduced as shown in Table 14.27.

Table 14.27: Matrix Further Reduced to 3×3 (2 Deleted Column)

		Player B		
		1	3	4
Player A	2	6	8	1
	3	8	15	1
	4	3	-1	4

Now, Row 2 is dominated by Row 3, hence delete Row 2, as shown in Table 14.28.

Table 14.28: Reduced Matrix (Row 2 Deleted)

		Player B		
		1	3	4
Player A	3	8	15	1
	4	3	-1	4

Now, as when comparing rows and columns, no column or row dominates the other. Since there is a tie while comparing the rows or columns, take the average of any two rows and compare. We have the following three combinations of matrices as shown in Table 14.29(a) (b) and (c).

Table 14.29 (a, b, c): Matrix Combinations

(a) B $\frac{R_1+R_3}{2} R_3$	(b) B $R_1 \frac{R_3+R_4}{2}$	(c) B $R_2 \frac{R_1+R_4}{2}$
$A \begin{pmatrix} 11.5 & 1 \\ 1 & 4 \end{pmatrix} \times$	$A \begin{pmatrix} 8 & 8 \\ 3 & 1.5 \end{pmatrix} \checkmark$	$A \begin{pmatrix} 15 & 4.5 \\ -1 & 3.5 \end{pmatrix} \times$

When comparing column 1 and the average of column 3 and column 4, column 1 is dominated by the average of column 3 and 4. Hence delete column 1. Finally, we get the 2×2 matrix as shown in Table 14.30.

Table 14.30: 2×2 Matrix After Deleting Column 1

		Player B	
		3	4
Player A	3	$\begin{pmatrix} 15 & 1 \\ -1 & 4 \end{pmatrix}$	
	4		

The strategy for the arrived matrix is a mixed strategy; using probability formula, we find p_1, p_2 and q_1, q_2 .

$$p_1 = \frac{4 - (-1)}{(15 + 4) - (1 + (-1))}$$

$$= \frac{5}{19}$$

$$p_2 = 1 - \frac{5}{9} = \frac{14}{9}$$

$$q_1 = \frac{4 - 1}{19} = \frac{3}{19}$$

$$q_2 = 1 - \frac{3}{19} = \frac{16}{19}$$

$$\text{Value of the game, } v = \frac{(15 \times 4) - (1 \times (-1))}{(15 + 4) - (1 + (-1))}$$

$$= \frac{60 + 1}{19}$$

$$= \frac{61}{19}$$

The optimum mixed strategies are given below in Table 14.31 (a, b)

Table 14.31(a) and (b): Optimum Mixed Strategies

$$(a) \quad S_A = \begin{pmatrix} A_1 & A_2 & A_3 & A_4 \\ 0 & 0 & \frac{5}{19} & \frac{14}{19} \end{pmatrix} \text{ and } (b) \quad S_B = \begin{pmatrix} B_1 & B_2 & B_3 & B_4 \\ 0 & 0 & \frac{3}{19} & \frac{16}{19} \end{pmatrix}$$

14.6 SOLVING PROBLEM ON THE COMPUTER WITH TORA

Pure strategy problem

Example 9: is solved using computer with TORA. From **Main menu** of TORA package select **Zero-sum Games** option. Click **Go to Input Screen** Enter the input values of the problem as shown in the Figure 14.1.

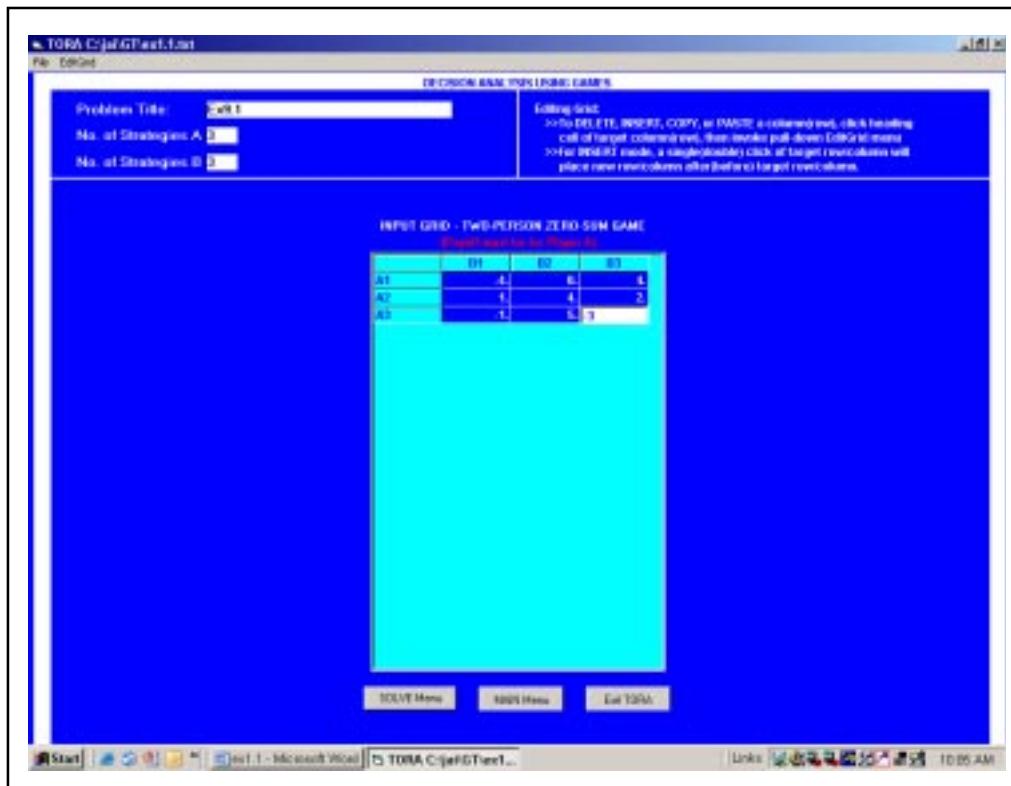


Figure 14.1: Solving Pure Strategy Problem Using TORA (Input Screen)

Now, go to Solve menu and click. Another screen appears with Solved Problem Select solve problem and click LP-based. Then select the output format screen and click Go to Output Screen. The following output screen is displayed, as shown in Figure 14.2.

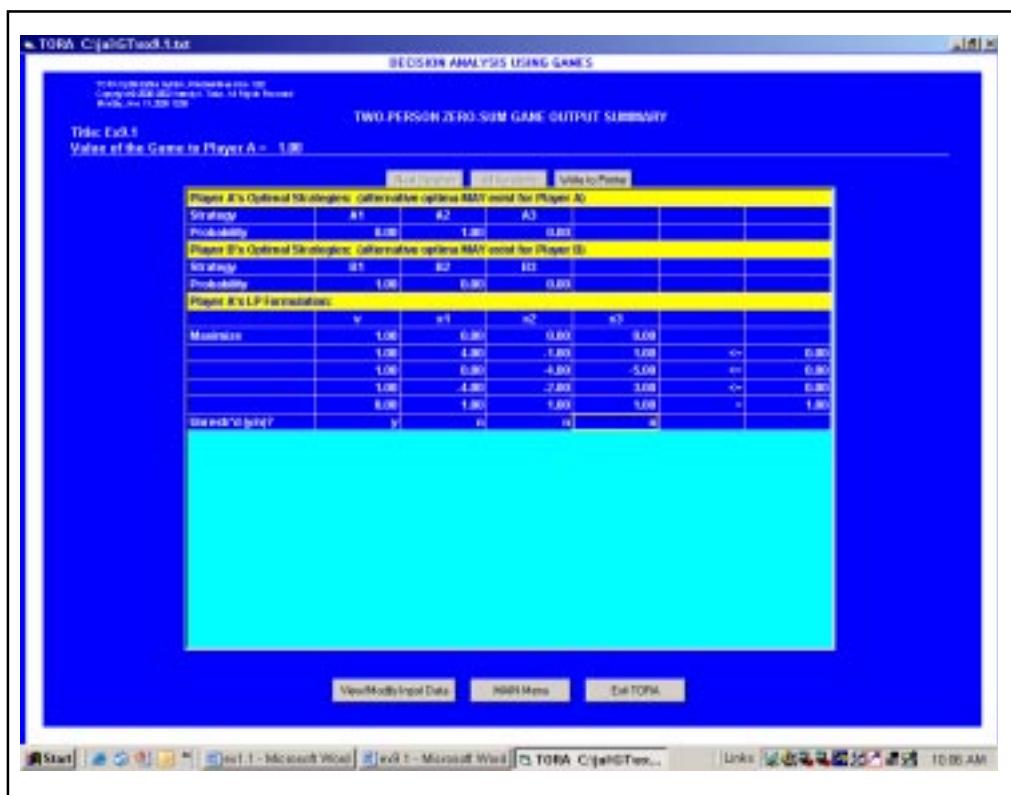


Figure 14.2: Solving Pure Strategy Problem Using TORA (Output Screen)

The results of the problem can be read directly from the output screen.

Value of the Game to Player A = 1.00

Player A optimal strategies:

Strategies:	A ₁	A ₂	A ₃
Probability:	0	1	0

Player B optimal strategies:

Strategies:	B ₁	B ₂	B ₃
Probability:	1	0	0

The output also includes the linear programming formulation for Player A.

Mixed Strategy Problem

Table 14.32

		Player B	
		1	2
Player A	1	5	2
	2	3	4

The output screen for the problem is shown in Figure 14.3

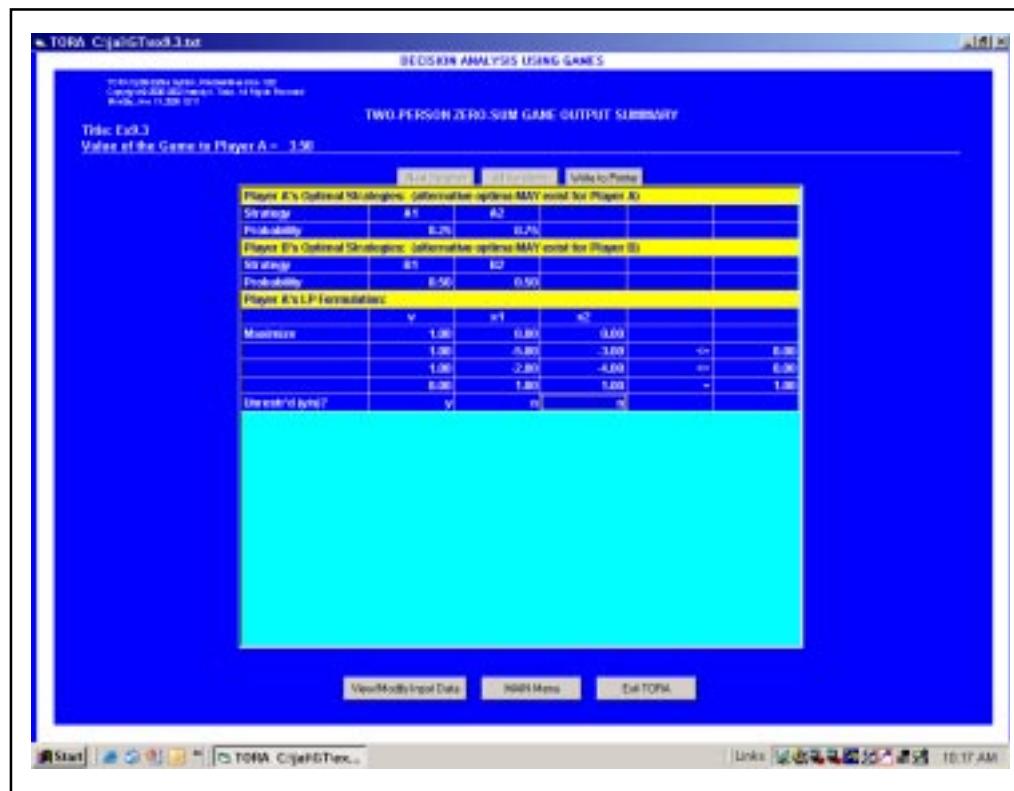


Figure 14.3: Solving Mixed Strategy Problems Using TORA (Output Screen)

Here the players play both the strategies in what turns out to be a mixed strategy game.

$$\begin{array}{ccccc} & \text{A}_1 & \text{A}_2 & & \\ \text{Player A :} & 0.25 & 0.75 & \text{Player B :} & 0.5 \quad 0.5 \end{array}$$

Value of the game, $v = 3.50$.

Example 10: Solve the following 2×3 game given below in Table 14.33 graphically, using computer.

Table 14.33: Game Problem

		Player B		
		B ₁	B ₂	B ₃
A ₁	A ₁	1	3	10
	A ₂	9	5	2

Solution: The game does not possess any saddle point and hence the solution has mixed strategies.

A's expected payoffs against B's pure moves are given by

Table 14.34: Mixed Strategies Compared

B's pure strategy	A's expected payoffs
B ₁	$p_1 + 9(1-p_1) = -8p_1 + 9$
B ₂	$3p_1 + 5(1-p_1) = -2p_1 + 5$
B ₃	$10p_1 + 2(1-p_1) = 8p_1 + 2$

The expected payoff equations are plotted as functions of p_1 which show the payoffs of each column represented as points on two vertical axis. Strategy B_1 is plotted by joining value 1 on axis 2 with the value 9 on axis 1. Similarly, other equations are drawn. The output using TORA is given in the Figure 14.4 below:

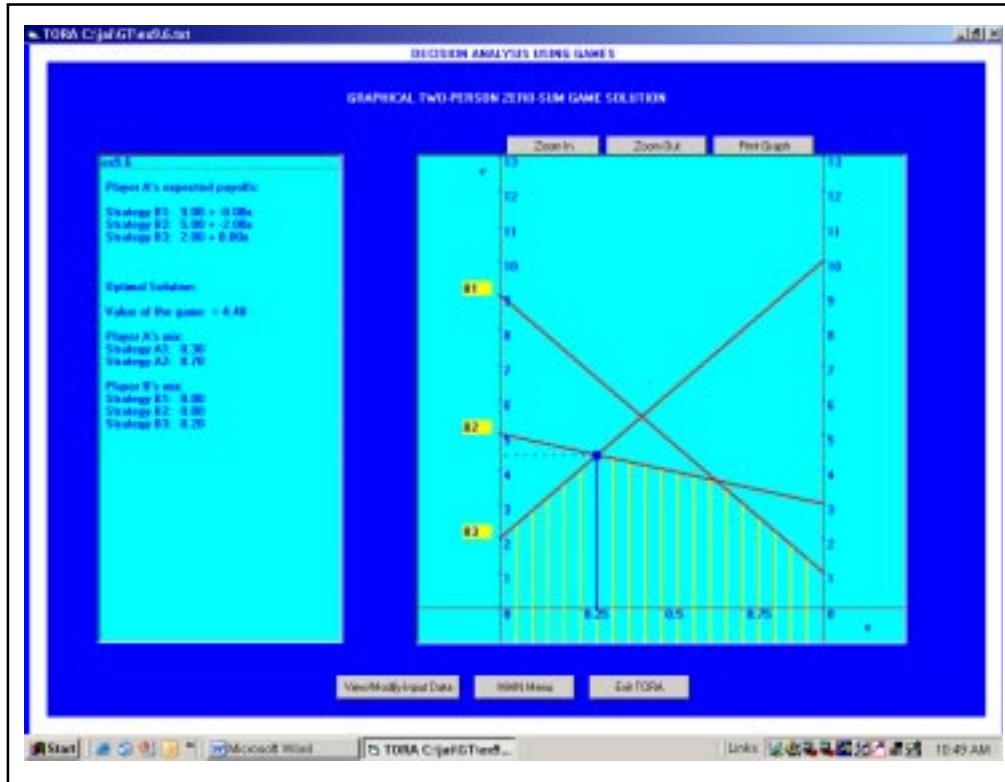


Figure 14.4: Graphical Solution of Game Using TORA (Output Screen)

Player A always wants to maximize his minimum expected payoff. Consider the highest point of intersection I on lower envelope of A's expected payoff equation. The lines B_2 and B_3 passing through I, are the strategies that B needs to play. Therefore the given matrix is reduced to 2×2 matrix as shown in Table 14.35.

Table 14.35: Reduced 2×2 Matrix

$$\begin{array}{cc} & B_2 & B_3 \\ A_1 & 3 & 10 \\ A_2 & 5 & 2 \end{array}$$

Solving the 2×2 matrix, the optimal strategies are obtained using the usual method

Table 14.36: Optimal Strategies

$$(a) S_A = \begin{pmatrix} A_1 & A_2 \\ 0.30 & 0.70 \end{pmatrix} \quad \text{and} \quad (b) S_B = \begin{pmatrix} B_1 & B_2 & B_3 \\ 0 & 0.80 & 0.20 \end{pmatrix}$$

The value of the game, $v = 4.40$.

14.7 SOLVING LP MODEL GAMES GRAPHICALLY USING COMPUTER

Example 11: Solve the following game shown in Table 14.37, by linear programming.

Table 14.37: Game Problem

		Player B		
		B ₁	B ₂	B ₃
Player A	A ₁	4	-1	-4
	A ₂	-3	4	-2

The linear programming formulation is given by,

For player A,

$$\text{Maximize, } z = v$$

Subject to the constraints,

$$v - 4x_1 + 3x_2 \leq 0 \quad \dots \dots \dots \text{(i)}$$

$$v + x_1 - 4x_2 \leq 0 \quad \dots \dots \dots \text{(ii)}$$

$$v + 4x_1 + x_2 \leq 0 \quad \dots \dots \dots \text{(iii)}$$

$$x_1 + x_2 = 1 \quad \dots \dots \dots \text{(iv)}$$

where, $x_1, x_2 \geq 0$

v is unrestricted.

For Player B,

$$\text{Minimize, } z = v$$

Subject to constraints,

$$v - 4y_1 + y_2 + 4y_3 \geq 0 \quad \dots \dots \dots \text{(v)}$$

$$v + 3y_1 - 4y_2 + y_3 \geq 0 \quad \dots \dots \dots \text{(vi)}$$

$$y_1 + y_2 + y_3 = 1 \quad \dots \dots \dots \text{(vii)}$$

where, $y_1, y_2, y_3 \geq 0$

v is unrestricted.

The problem can be solved by using linear programming. This can also be solved by using two-person zero-sum game. The output result is given in Figure 14.5 below:



Figure 14.5: Two-person Zero-sum Game, Output Result Using TORA

The optimal strategies are,

Game Theory

	A ₁	A ₂
Player A :	0.11	0.89
	B ₁	B ₂
Player B :	0.22	0 0.78

Value of the game, v = - 2.22

Check Your Progress 14.2

Take a type of business problem of your choice in which game theory will be helpful.

- Notes:**
- (a) Write your answer in the space given below.
 - (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
 - (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-
-

14.8 LET US SUM UP

At the outset, a comprehensive overview of games theory has been applicable to those competitive situation which are technically known as “competitive games”. If we elaborately go into the depth of the word “competitions” it is a watchword of contemporary life and competitive situation exists if two or more individually are making decisions in a circumstances that involves conflicting interests and in which the result is controlled by the decision of all parties concerned. The initial discussion study with the zero person game and moves to saddle point further principle of dominance applicable to reduce the size of matrix and their is a $2 \times n$ of $M \times 2$ games which can be solved with the help of graphical method. In short, the game theory facilitates us to learn how to approves and comprehend a conflict circumstance and to enhance the decision procedure.

14.9 LESSON-END ACTIVITY

Apply the game theory approach of two television vendor line LG & Samsung for their advertising strategy.

14.10 KEYWORDS

Two Person Game : A game that only has two players.

Zero Sum Game : A game in which one player wins and other player loses.

Dominance : A process by which the size of the game will be reduced.

Strategy	: The strategy of a player is the list of all possible actions that he takes for every pay-off. The strategy is classified into pure strategy and mixed strategy.
Pure Strategy	: Pure strategy is always selecting a particular course of action with the probability of 1. For example, in case of two strategies, probability of selecting the strategies for players A is $p_1 = 0$ and $p_2 = 1$.
Mixed Strategy	: Mixed strategy is to choose at least two courses of action. The probability of selecting an individual strategy will be less than 1, but the sum of the strategies will be 1. For example, if player A plays a mixed strategy, then the probability of selection of mixed strategy is $p_1 = 0.45$ and $p_2 = 0.55$. But the sum of the strategies is $0.45 + 0.55 = 1$.
Saddle Point	: Saddle point is a situation where both the players are facing pure strategies. When there is no saddle point, it indicates the players will play both the strategies.
Minimax Criterion	: Minimax criterion is selecting the strategies that minimize the loss for each player. In other words, the player always anticipates worst possible outcome and chooses the strategy to get maximum for profit and minimum for loss.
Value of the Game	: The Value of the game is the expected gain of player A if both players use their best strategies. The best strategy is arrived at using minimax criterion.

14.11 QUESTIONS FOR DISCUSSION

1. **Write True or False against each statement:**
 - (a) Graphical method can only be used in games with no saddle point.
 - (b) Concept of dominance is very useful for expanding the size of the matrix.
 - (c) Saddle point in a pay off matrix is one which is smallest value in its row and the largest value in its column.
 - (d) In two-person zero-sum game there will be more than two choices.
2. **Briefly comment on the following:**
 - (a) Dominance occurs in the pay-off matrix.
 - (b) Best strategic are mixed strategies if there is no involvement of saddle point.
 - (c) Graphical method is feasible for Small values.
 - (d) When the game have no saddle point & also cannot be reduced by dominance.
 - (e) In game theory we determine the best strategies for each player.
 - (f) A saddle point is an element of the matrix.
3. **Fill in the blank:**
 - (a) Game theory applies to those _____ situation which are technically known as “competitive game”.
 - (b) Strategy could be _____ or one.
 - (c) A game involving n-players is called a _____ game.
 - (d) Every course of action is a _____ strategy.
 - (e) In game theory all players act _____.

4. Write short Notes on following:

Game Theory

- (a) The value of a game
- (b) The sum & non-zero-sum games.
- (c) Maximum & Minimum strategy
- (d) Concept of dominance.
- (e) Pure strategy
- (f) Mixed strategy
- (g) Pay-off matrix
- (h) Saddle point
- (i) Optimum strategies.

14.12 TERMINAL QUESTIONS

1. What are the properties of Two-person Zero-sum game?
2. Define Pure strategy and Mixed strategy.
3. What is meant by the saddle point?
4. What is meant by a Fair Game?
5. Explain how games can be solved using the dominance property.

Exercise Problems

1. Using maximin criteria, identify whether the players play pure strategy or mixed strategies

		Player B	
		1	2
1	1	-7	3
	2	-5	-2

		Player B	
		1	2
1	1	7	2
	2	1	5

2. Solve the game and determine whether it is strictly determinable.

		Player B		
		B ₁	B ₂	B ₃
Player A	A ₁	3	-1	2
	A ₂	-2	5	7
	A ₃	2	3	-5

(b)

		Player B			
		B ₁	B ₂	B ₃	
Player A		A ₁	6	10	8
		A ₂	-4	10	14
		A ₃	4	6	-10

3. Determine the optimum strategies of players.

		B				B				
		1	2	3	4	1	2	3	4	
A		1	10	-30	20	-40	1	-1	2	-2
		2	5	-5	10	15	2	3	-2	0
		3	20	20	-30	10	3	4	-2	3
		4	5	-10	20	15	4	0	1	3

4. Solve the game.

		Company B			Company B				
		B ₁	B ₂	B ₃	B ₁	B ₂	B ₃		
Company A		A ₁	15	25	35	A ₁	7	5	2
		A ₂	5	10	45	A ₂	-1	3	-2
		A ₃	65	55	35	A ₃	4	3	-7

5. Consider the payoff matrix of player A and solve.

		Player B						
		1	2	3	4	5	6	
Player A		1	6	3	7	-4	6	8
		2	7	-11	8	4	7	9

6. Solve the following sequence game using dominance property.

		Company A				
		I	II	III	IV	
Company B		A	14	4	8	18
		B	8	3	2	12
		C	8	7	-6	16
		D	6	5	12	10

7. Use dominance property to solve the following game.

	B ₁	B ₂	B ₃	B ₄	B ₅
A ₁	4	4	2	-4	-6
A ₂	8	6	8	-4	0
A ₃	10	2	4	0	12

8. Solve the following two-person zero-sum game to find the value of the game.

		Company B			
		1	2	3	4
Company A	1	2	-2	4	1
	2	6	1	12	3
	3	-3	2	0	6
	4	2	-3	7	7

9. Solve the game whose payoff matrix is given for a 2×2 matrix.

(a)	$\begin{pmatrix} 5 & -2 \\ -1 & 7 \end{pmatrix}$	(b)	$\begin{pmatrix} 2 & 5 \\ 2 & -2 \end{pmatrix}$
-----	--	-----	---

10. Solve the game graphically.

		B			
		B ₁	B ₂	B ₃	B ₄
A	A ₁	4	-2	3	-1
	A ₂	-1	2	0	1

11. Use dominance property to reduce the matrix and solve it graphically

		Player B			
		A	B	C	D
Player A	1	18	4	6	4
	2	6	2	13	7
	3	11	5	17	3
	4	7	6	12	2

12. Formulate a linear programming model for the following game.

		Player B				
		1	2	3	4	5
Player A	1	6	15	30	21	6
	2	3	3	6	6	4
	3	12	12	24	36	3

14.13 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

1. (a) True (b) False (c) True (d) False
3. (a) competitive (b) pure, mixed (c) n-person (d) pure
(e) intelligently.

14.14 SUGGESTED READINGS

Davis. M, *Game Theory: A Nontechnical Introduction*, New York; Basic Books, Inc.

Lucas, "An Overview of the Mathematical Theory of Game" Management Science Vol. 8, No. 8, Part-II.

Luce R.D. & Raiffa. H, *Game & Decisions*, New York, John Wiley & Sam.

Rapport, A., *Two Person Game Theory*, Anne Arbric, Michiger: The University of Michigan Press, 1966.

Shubik. M, *The Uses & Methods of Game Theory*, New York; American Elsevier.

Van Neumann, J. & Morgen Stan. O, *Theory of Games & Economic Behaviour*.

LESSON

15

SIMULATION

CONTENTS

- 15.0 Aims and Objectives
- 15.1 Introduction
- 15.2 Advantages and Disadvantages of Simulation
- 15.3 Monte Carlo Simulation
- 15.4 Simulation of Demand Forecasting Problem
- 15.5 Simulation of Queuing Problems
- 15.6 Simulation of Inventory Problems
- 15.7 Let us Sum Up
- 15.8 Lesson-end Activities
- 15.9 Keywords
- 15.10 Questions for Discussion
- 15.11 Terminal Questions
- 15.12 Model Answers to Questions for Discussion
- 15.13 Suggested Readings

15.0 AIMS AND OBJECTIVES

This is the last lesson of the QT which will discuss about the Mathematical analysis and mathematical technique simulation technique is considered as a valuable tool because wide area of applications.

15.1 INTRODUCTION

In the previous chapters, we formulated and analyzed various models on real-life problems. All the models were used with mathematical techniques to have analytical solutions. In certain cases, it might not be possible to formulate the entire problem or solve it through mathematical models. In such cases, simulation proves to be the most suitable method, which offers a near-optimal solution. Simulation is a reflection of a real system, representing the characteristics and behaviour within a given set of conditions.

In simulation, the problem must be defined first. Secondly, the variables of the model are introduced with logical relationship among them. Then a suitable model is constructed. After developing a desired model, each alternative is evaluated by generating a series of values of the random variable, and the behaviour of the system is observed. Lastly, the results are examined and the best alternative is selected the whole process has been summarized and shown with the help of a flow chart in the Figure 90.

Simulation technique is considered as a valuable tool because of its wide area of application. It can be used to solve and analyze large and complex real world problems. Simulation provides solutions to various problems in functional areas like production, marketing, finance, human resource, etc., and is useful in policy decisions through corporate planning models. Simulation experiments generate large amounts of data and information using a small sample data, which considerably reduces the amount of cost and time involved in the exercise.

For example, if a study has to be carried out to determine the arrival rate of customers at a ticket booking counter, the data can be generated within a short span of time can be used with the help of a computer.

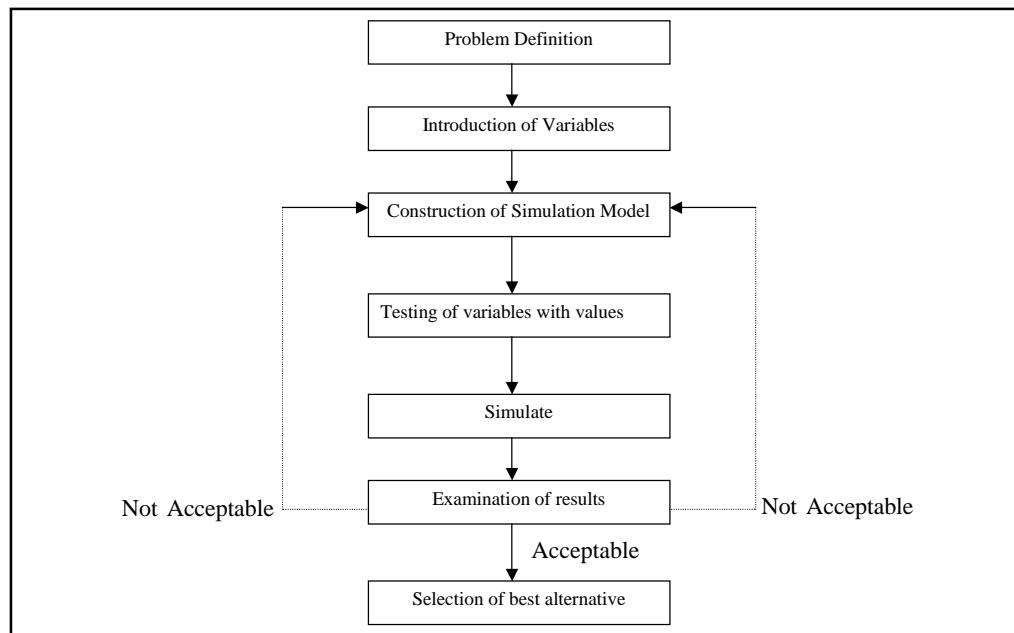


Figure 15.1: Simulation Process

15.2 ADVANTAGES AND DISADVANTAGES OF SIMULATION

Advantages

- Simulation is best suited to analyze complex and large practical problems when it is not possible to solve them through a mathematical method.
- Simulation is flexible, hence changes in the system variables can be made to select the best solution among the various alternatives.
- In simulation, the experiments are carried out with the model without disturbing the system.
- Policy decisions can be made much faster by knowing the options well in advance and by reducing the risk of experimenting in the real system.

Disadvantages

- Simulation does not generate optimal solutions.
- It may take a long time to develop a good simulation model.
- In certain cases simulation models can be very expensive.
- The decision-maker must provide all information (depending on the model) about the constraints and conditions for examination, as simulation does not give the answers by itself.

15.3 MONTE CARLO SIMULATION

In simulation, we have deterministic models and probabilistic models. Deterministic simulation models have the alternatives clearly known in advance and the choice is made by considering the various well-defined alternatives. Probabilistic simulation model is stochastic in nature and all decisions are made under uncertainty. One of the probabilistic simulation models is the Monte Carlo method. In this method, the decision variables are represented by a probabilistic distribution and random samples are drawn from probability distribution using random numbers. The simulation experiment is conducted until the required number of simulations are generated. Finally, the best course of action is selected for implementation. The significance of Monte Carlo Simulation is that decision variables may not explicitly follow any standard probability distribution such as Normal, Poisson, Exponential, etc. The distribution can be obtained by direct observation or from past records.

Procedure for Monte Carlo Simulation:

Step 1: Establish a probability distribution for the variables to be analyzed.

Step 2: Find the cumulative probability distribution for each variable.

Step 3: Set Random Number intervals for variables and generate random numbers.

Step 4: Simulate the experiment by selecting random numbers from random numbers tables until the required number of simulations are generated.

Step 5: Examine the results and validate the model.

15.4 SIMULATION OF DEMAND FORECASTING PROBLEM

Example 1: An ice-cream parlor's record of previous month's sale of a particular variety of ice cream as follows (see Table 15.1).

Table 15.1: Simulation of Demand Problem

Demand (No. of Ice-creams)	No. of days
4	5
5	10
6	6
7	8
8	1

Simulate the demand for first 10 days of the month

Solution: Find the probability distribution of demand by expressing the frequencies in terms of proportion. Divide each value by 30. The demand per day has the following distribution as shown in Table 15.2.

Table 15.2: Probability Distribution of Demand

Demand	Probability
4	0.17
5	0.33
6	0.20
7	0.27
8	0.03

Find the cumulative probability and assign a set of random number intervals to various demand levels. The probability figures are in two digits, hence we use two digit random numbers taken from a random number table. The random numbers are selected from the table from any row or column, but in a consecutive manner and random intervals are set using the cumulative probability distribution as shown in Table 15.3.

Table 15.3: Cumulative Probability Distribution

Demand	Probability	Cumulative Probability	Random Number Interval
4	0.17	0.17	00-16
5	0.33	0.50	17-49
6	0.20	0.70	50-69
7	0.27	0.97	70-96
8	0.03	1.00	97-99

To simulate the demand for ten days, select ten random numbers from random number tables. The random numbers selected are,

17, 46, 85, 09, 50, 58, 04, 77, 69 and 74

The first random number selected, 7 lies between the random number interval 17-49 corresponding to a demand of 5 ice-creams per day. Hence, the demand for day one is 5. Similarly, the demand for the remaining days is simulated as shown in Table 15.4.

Table 15.4: Demand Simulation

Day	1	2	3	4	5	6	7	8	9	10
Random Number	17	46	85	09	50	58	04	77	69	74
Demand	5	5	7	4	6	6	4	7	6	7

Example 2: A dealer sells a particular model of washing machine for which the probability distribution of daily demand is as given in Table 15.5.

Table 15.5: Probability Distribution of Daily Demand

Demand/day	-	0	1	2	3	4	5
Demand	-	0.05	0.25	0.20	0.25	0.10	0.15

Find the average demand of washing machines per day.

Solution: Assign sets of two digit random numbers to demand levels as shown in Table 15.6.

Table 15.6: Random Numbers Assigned to Demand

Demand	Probability	Cumulative Probability	Random Number Intervals
0	0.05	0.05	00-04
1	0.25	0.30	05-29
2	0.20	0.50	30-49
3	0.25	0.75	50-74
4	0.10	0.85	75-84
5	0.15	1.00	85-99

Ten random numbers that have been selected from random number tables are 68, 47, 92, 76, 86, 46, 16, 28, 35, 54. To find the demand for ten days see the Table 15.7 below.

Table 15.7: Ten Random Numbers Selected

Trial No	Random Number	Demand / day
1	68	3
2	47	2
3	92	5
4	76	4
5	86	5
6	46	2
7	16	1
8	28	1
9	35	2
10	54	3
Total Demand		28

Average demand = $28/10 = 2.8$ washing machines per day.

The expected demand /day can be computed as,

where, p_i = probability and x_i = demand

$$= (0.05 \times 0) + (0.25 \times 1) + (0.20 \times 2) + (0.25 \times 3) + (0.1 \times 4) + (0.15 \times 5)$$

= 2.55 washing machines.

The average demand of 2.8 washing machines using ten-day simulation differs significantly when compared to the expected daily demand. If the simulation is repeated number of times, the answer would get closer to the expected daily demand.

Example 3: A farmer has 10 acres of agricultural land and is cultivating tomatoes on the entire land. Due to fluctuation in water availability, the yield per acre differs. The probability distribution yields are given below:

- a. The farmer is interested to know the yield for the next 12 months if the same water availability exists. Simulate the average yield using the following random numbers 50, 28, 68, 36, 90, 62, 27, 50, 18, 36, 61 and 21, given in Table 15.8.

Table 15.8: Simulation Problem

Yield of tomatoes per acre (kg)	Probability
200	0.15
220	0.25
240	0.35
260	0.13
280	0.12

- b. Due to fluctuating market price, the price per kg of tomatoes varies from Rs. 5.00 to Rs. 10.00 per kg. The probability of price variations is given in the Table 216 below. Simulate the price for next 12 months to determine the revenue per acre. Also find the average revenue per acre. Use the following random numbers 53, 74, 05, 71, 06, 49, 11, 13, 62, 69, 85 and 69.

Table 15.9: Simulation Problem

Price per kg (Rs)	Probability
5.50	0.05
6.50	0.15
7.50	0.30
8.00	0.25
10.00	0.15

Solution:

Table 15.10: Table for Random Number Interval for Yield

Yield of tomatoes per acre	Probability	Cumulative Probability	Random Number Interval
200	0.15	0.15	00 – 14
220	0.25	0.40	15 – 39
240	0.35	0.75	40 – 74
260	0.13	0.88	75 – 87
280	0.12	1.00	88 – 99

Table 15.11: Table for Random Number Interval for Price

Price Per Kg	Probability	Cumulative Probability	Random Number Interval
5.00	0.05	0.05	00 – 04
6.50	0.15	0.20	05 – 19
7.50	0.30	0.50	20 – 49
8.00	0.25	0.75	50 – 74
10.00	0.25	1.00	75 – 99

Table 15.12: Simulation for 12 months period

Month (1)	Yield (2)	Price (3)	Revenue / Acre (4) = 2 × 3 (Rs)
1	240	8.00	1960
2	220	8.00	1760
3	240	6.50	1560
4	220	8.00	1760
5	250	6.50	1820
6	240	7.50	1800
7	220	6.50	1430
8	240	6.50	1560
9	220	8.00	1760
10	220	8.00	1760
11	240	10.00	2400
12	220	8.00	1760

$$\text{Average revenue per acre} = 21330 / 12$$

$$= \text{Rs. } 1777.50$$

Example 4: J.M Bakers has to supply only 200 pizzas every day to their outlet situated in city bazaar. The production of pizzas varies due to the availability of raw materials and labor for which the probability distribution of production by observation made is as follows:

Table 15.13: Simulation Problem

Production per day	196	197	198	199	200	201	202	203	204
Probability	0.06	0.09	0.10	0.16	0.20	0.21	0.08	0.07	0.03

Simulate and find the average number of pizzas produced more than the requirement and the average number of shortage of pizzas supplied to the outlet.

Solution: Assign two digit random numbers to the demand levels as shown in Table 15.14

Table 15.14: Random Numbers Assigned to the Demand Levels

Demand	Probability	Cumulative Probability	No of Pizzas shortage
196	0.06	0.06	00-05
197	0.09	0.15	06-14
198	0.10	0.25	15-24
199	0.16	0.41	25-40
200	0.20	0.61	41-60
201	0.21	0.82	61-81
202	0.08	0.90	82-89
203	0.07	0.97	90-96
204	0.03	1.00	97-99

Selecting 15 random numbers from random numbers table and simulate the production per day as shown in Table 15.15 below.

Simulation

Table 15.15: Simulation of Production Per Day

Trial Number	Random Number	Production Per day	No of Pizzas over produced	No of pizzas shortage
1	26	199	-	1
2	45	200	-	-
3	74	201	1	-
4	77	201	1	-
5	74	201	1	-
6	51	200	-	-
7	92	203	3	-
8	43	200	-	-
9	37	199	-	1
10	29	199	-	1
11	65	201	1	-
12	39	199	-	1
13	45	200	-	-
14	95	203	3	-
15	93	203	3	-
		Total	12	4

The average number of pizzas produced more than requirement

$$= 12/15$$

$$= 0.8 \text{ per day}$$

The average number of shortage of pizzas supplied

$$= 4/15$$

$$= 0.26 \text{ per day}$$

Check Your Progress 15.1

1. Discuss the role of simulation in demand forecasting.
2. What is Monte Carlo simulation?

- Notes:** (a) Write your answer in the space given below.
- (b) Please go through the lesson sub-head thoroughly you will get your answers in it.
- (c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.
-
-
-
-

15.5 SIMULATION OF QUEUING PROBLEMS

Example 5: Mr. Srinivasan, owner of Citizens restaurant is thinking of introducing separate coffee shop facility in his restaurant. The manager plans for one service counter for the coffee shop customers. A market study has projected the inter-arrival times at the restaurant as given in the Table 15.16. The counter can service the customers at the following rate:

Table 15.16: Simulation of Queuing Problem

Inter-arrival times		Service times	
Time between two consecutive arrivals (minutes)	Probability	Service time (minutes)	Probability
2	0.15	2	0.10
3	0.25	3	0.25
4	0.20	4	0.30
5	0.25	5	0.2
6	0.15	6	0.15

Mr. Srinivasan will implement the plan if the average waiting time of a customers in the system is less than 5 minutes.

Before implementing the plan, Mr. Srinivasan would like to know the following:

- i. Mean waiting time of customers, before service.
- ii. Average service time.
- iii. Average idle time of service.
- iv. The time spent by the customer in the system.

Simulate the operation of the facility for customer arriving sample of 20 cars when the restaurant starts at 7.00 pm every day and find whether Mr. Srinivasan will go for the plan.

Solution: Allot the random numbers to various inter-arrival service times as shown in Table 15.17.

Table 15.17: Random Numbers Allocated to Various Inter-Arrival Service Times

Sl. No.	Random Number (Arrival)	Inter Arrival Time (Min)	Arrival Time at	Service Starts at	Random Number (service)	Service Time (Min)	Service Ends at	Waiting Time	
								Customer	Service (Min)
1	87	6	7.06	7.06	36	4	7.10	-	6
2	37	3	7.09	7.10	16	3	7.13	1	-
3	92	6	7.15	7.15	81	5	7.20	-	2
4	52	4	7.19	7.20	08	2	7.22	1	-
5	41	4	7.23	7.23	51	4	7.27	-	1
6	05	2	7.25	7.27	34	3	7.30	2	-
7	56	4	7.29	7.30	88	6	7.36	1	-
8	70	5	7.34	7.36	88	6	7.42	2	-
9	70	5	7.39	7.42	15	3	7.45	3	-
10	07	2	7.41	7.45	53	4	7.49	4	-
11	86	6	7.47	7.49	01	2	7.51	2	-
12	74	5	7.52	7.52	54	4	7.56	-	1
13	31	3	7.55	7.56	03	2	7.58	1	-
14	71	5	8.00	8.00	54	4	8.04	1	2
15	57	4	8.04	8.04	56	4	8.08	-	-
16	85	6	8.10	8.10	05	2	8.12	-	2
17	39	3	8.13	8.13	01	2	8.15	-	1
18	41	4	8.17	8.17	45	4	8.21	-	2
19	18	3	8.20	8.21	11	3	8.24	1	-
20	38	3	8.23	8.24	76	5	8.29	1	-
	Total	83				72		20	17

- i. Mean waiting time of customer before service = $20/20 = 1$ minute
- ii. Average service idle time = $17/20 = 0.85$ minutes
- iii. Time spent by the customer in the system = $3.6 + 1 = 4.6$ minutes.

Example 6: Dr. Strong, a dentist schedules all his patients for 30 minute appointments. Some of the patients take more or less than 30 minutes depending on the type of dental work to be done. The following Table 15.18 shows the summary of the various categories of work, their probabilities and the time actually needed to complete the work.

Table 15.18: Simulation Problem

Category	Time required (minutes)	Probability of category
Filling	45	0.40
Crown	60	0.15
Cleaning	15	0.15
Extraction	45	0.10
Check-up	15	0.20

Simulate the dentist's clinic for four hours and determine the average waiting time for the patients as well as the idleness of the doctor. Assume that all the patients show up at the clinic exactly at their scheduled arrival time, starting at 8.00 am. Use the following random numbers for handling the above problem: 40,82,11,34,25,66,17,79.

Solution: Assign the random number intervals to the various categories of work as shown in Table 15.19.

Table 15.19: Random Number Intervals Assigned to the Various Categories

Category of work	Probability	Cumulative probability	Random Number Interval
Filling	0.40	0.40	00-39
Crown	0.15	0.55	40-54
Cleaning	0.15	0.70	55-69
Extraction	0.10	0.80	70-79
Check-up	0.20	1.00	80-99

Assuming the dentist clinic starts at 8.00 am, the arrival pattern and the service category are shown in Table 15.20.

Table 15.20: Arrival Pattern of the Patients

Patient Number	Scheduled Arrival	Random Number	Service category	Service Time
1	8.00	40	Crown	60
2	8.30	82	Check-up	15
3	9.00	11	Filling	45
4	9.30	34	Filling	45
5	10.00	25	Filling	45
6	10.30	66	Cleaning	15
7	11.00	17	Filling	45
8	11.30	79	Extraction	45

Table 15.21: The arrival, departure patterns and patients' waiting time are tabulated.

Time	Event (Patient Number)	Patient Number (Time to go)	Waiting (Patient Number)
8.00	1 arrives	1 (60)	-
8.30	2 arrives	1 (30)	2
9.00	1 departure, 3 arrives	2 (15)	3
9.15	2 depart	3 (45)	-
9.30	4 arrive	3 (30)	4
10.00	3 depart, 5 arrive	4 (45)	5
10.30	6 arrive	4 (15)	5,6
10.45	4 depart	5 (45)	6
11.00	7 arrive	5 (30)	6,7
11.30	5 depart, 8 arrive	6 (15)	7,8
11.45	6 depart	7 (45)	8
12.00	End	7 (30)	8

The dentist was not idle during the simulation period. The waiting times for the patients are as given in Table 15.22 below.

Table 15.22: Patient's Waiting Time

Patient	Arrival Time	Service Starts	Waiting time (minutes)
1	8.00	8.00	0
2	8.30	9.00	30
3	9.00	9.15	15
4	9.30	10.00	30
5	10.00	10.45	45
6	10.30	11.30	60
7	11.00	11.45	45
8	11.30	12.30	60
		Total	285

The average waiting time of patients = $285/8$

$$= 35.625 \text{ minutes.}$$

15.6 SIMULATION OF INVENTORY PROBLEMS

A dealer of electrical appliances has a certain product for which the probability distribution of demand per day and the probability distribution of the lead-time, developed by past records are as shown in Table 15.23 and 10.24 respectively

Table 15.23: Probability distribution of lead demand

Demand (Units)	2	3	4	5	6	7	8	9	10
Probability	0.05	0.07	0.09	0.15	0.20	0.21	0.10	0.07	0.06

Table 15.24: Probability distribution of lead time

Lead Time (Days)	1	2	3	4
Probability	0.20	0.30	0.35	0.15

The various costs involved are,

Ordering Cost = Rs. 50 per order

Holding Cost = Rs. 1 per unit per day

Shortage Cost = Rs. 20 per unit per day

The dealer is interested in having an inventory policy with two parameters, the reorder point and the order quantity, i.e., at what level of existing inventory should an order be placed and the number of units to be ordered. Evaluate a simulation plan for 35 days, which calls for a reorder quantity of 35 units and a re-order level of 20 units, with a beginning inventory balance of 45 units.

Solution: Assigning of random number intervals for the demand distribution and lead-time distribution is shown in Tables 15.25 and 15.26 respectively.

Table 15.25: Random Numbers Assigned for Demand Per Day

Demand per day	Probability	Cumulative probability	Random Number Interval
2	0.05	0.05	00-04
3	0.07	0.12	05-11
4	0.09	0.21	12-20
5	0.15	0.36	21-35
6	0.20	0.56	36-55
7	0.21	0.77	56-76
8	0.10	0.87	77-86
9	0.07	0.94	87-93
10	0.06	1.00	94-99

Table 15.26: Random Numbers Assigned for Lead-time

Simulation

Lead Time (Days)	Probability	Cumulative probability	Random Number Interval
1	0.20	0.20	00-19
2	0.30	0.50	20-49
3	0.35	0.85	50-84
4	0.15	1.00	85-99

Table 15.27: Simulation Work-sheet for Inventory Problem (Case – 1)

Reorder Quantity = 35 units, Reorder Level = 20 units, Beginning Inventory = 45 units

Day	Random Number (Demand)	Demand	Random Number (Lead Time)	Lead Time (Days)	Inventory at end of day	Qty. Received	Ordering Cost	Holding Cost	Shortage Cost
0	-	-	-	-	45	-	-	-	-
1	58	7	-	-	38	-	-	38	-
2	45	6	-	-	32	-	-	32	-
3	43	6	-	-	26	-	-	26	-
4	36	6	73	3	20	-	50	20	-
5	46	6	-	-	14	-	-	14	-
6	46	6	-	-	8	-	-	8	-
7	70	7	-	-	1	35	-	36	-
8	32	5	-	-	31	-	-	31	-
9	12	4	-	-	27	-	-	27	-
10	40	6	-	-	21	-	-	21	-
11	51	6	21	2	15	-	50	15	-
12	59	7	-	-	8	-	-	8	-
13	54	6	-	-	37	35	-	37	-
14	16	4	-	-	33	-	-	33	-
15	68	7	-	-	26	-	-	26	-
16	45	6	45	2	20	-	50	20	-
17	96	10	-	-	10	-	-	10	-
18	33	5	-	-	40	35	-	40	-
19	83	8	-	-	32	-	-	32	-
20	77	8	-	-	24	-	-	24	-
21	05	3	-	-	21	-	-	21	-
22	15	4	76	3	17	-	50	17	-
23	40	6	-	-	11	-	-	11	-
24	43	6	-	-	5	-	-	5	-
25	34	5	-	-	35	35	-	35	-
26	44	6	-	-	29	-	-	29	-
27	89	9	96	4	20	-	50	20	-
28	20	4	-	-	16	-	-	16	-
29	69	7	-	-	9	-	-	9	-
30	31	5	-	-	4	-	-	4	-
31	97	10	-	-	29	35	-	29	-
32	05	3	-	-	26	-	-	26	-
33	59	7	94	4	19	-	50	19	-
34	02	2	-	-	17	-	-	17	-
35	35	5	-	-	12	-	-	12	-
					Total	300	768	-	

Table 15.28: Simulation Work-sheet for Inventory Problem (Case – II)

Reorder Quantity = 30 units, Reorder Level = 20 units, Beginning Inventory = 45 units

Day	Random Number (Demand)	Demand	Random Number (Lead Time)	Lead Time (Days)	Inventory at end of day	Qty. Received	Ordering Cost	Holding Cost	Shortage Cost
0	-	-	-	-	45	-	-	-	-
1	58	7	-	-	38	-	-	38	-
2	45	6	-	-	32	-	-	32	-
3	43	6	-	-	26	-	-	26	-
4	36	6	73	3	20	-	50	20	-
5	46	6	-	-	14	-	-	14	-
6	46	6	-	-	8	-	-	8	-
7	70	7	-	-	31	30	-	31	-
8	32	5	-	-	29	-	-	29	-
9	12	4	-	-	25	-	-	25	-
10	40	6	-	-	19	-	50	19	-
11	51	6	21	2	13	-	-	13	-
12	59	7	-	-	38	-	-	38	-
13	54	6	-	-	32	30	-	32	-
14	16	4	-	-	21	-	-	21	-
15	68	7	-	-	21	-	-	21	-
16	45	6	45	2	15	-	50	15	-
17	96	10	-	-	5	-	-	5	-
18	33	5	-	-	30	-	-	30	-
19	83	8	-	-	22	-	-	22	-
20	77	8	-	-	14	-	50	14	-
21	05	3	-	-	11	-	-	11	-
22	15	4	76	3	7	-	-	7	-
23	40	6	-	-	31	30	-	31	-
24	43	6	-	-	14	-	-	14	-
25	34	5	-	-	20	-	50	20	-
26	44	6	-	-	14	-	-	14	-
27	89	9	96	4	5	-	-	5	-
28	20	4	-	-	1	-	-	1	-
29	69	7	-	-	24	30	-	24	-
30	31	5	-	-	19	-	50	19	-
31	97	10	-	-	9	-	-	9	-
32	05	3	-	-	6	-	-	6	-
33	59	7	94	4	0	-	-	-	20
34	02	2	-	-	28	30	-	28	-
35	35	5	-	-	23	-	-	23	-
						Total	300	683	20

The simulation of 35 days with an inventory policy of reordering quantity of 35 units at the time of inventory level at the end of day is 20 units, as worked out in Table 10.27. The table explains the demand, inventory level, quantity received, ordering cost, holding cost and shortage cost for each day.

Completing a 35 day period, the costs are

Simulation

$$\text{Total ordering cost} = (6 \times 50) = \text{Rs } 300.00$$

$$\text{Total holding cost} = \text{Rs. } 768.00$$

Since the demand for each day is satisfied, there is no shortage cost.

Therefore, Total cost = $300 + 768$

$$= \text{Rs. } 1068.00$$

For a different set of parameters, with a re-order quantity of 30 units and the same re-order level of 20 units, if the 35-day simulation is performed, we get the total of various costs as shown in Table 10.28.

$$\text{Total ordering cost} = 6 \times 50 = \text{Rs. } 300.00$$

$$\text{Total holding cost} = \text{Rs. } 683.0$$

$$\text{Total shortage cost} = \text{Rs. } 20.00$$

Therefore,

$$\text{Total cost} = 300 + 683 + 20$$

$$= \text{Rs. } 1003.00$$

If we analyze the combination of both the parameters, Case II has lesser total cost than Case I. But at the same time, it does not satisfy the demand on 33rd day, that might cause customer dissatisfaction which may lead to some cost.

In this type of problems, the approach with various combinations of two parameter values is simulated a large number of times to find the total cost of each experiment, compare the total cost and select the optimum alternative, i.e., that one which incurs the lowest cost.

Check Your Progress 15.2

1. Explain how computer make ideal aides in simulating complex tasks.
2. What are the two types of computer programming languages that are available to facilitate the simulation process?
3. Why is the computer necessary in conducting a real world simulation.
4. Do you think the application of simulation will enhance strongly in the coming 10 years.
5. Draw a flow diagram for the simulation of electric-maintenance by the power corporation of India Ltd.

Notes: (a) Write your answer in the space given below.

(b) Please go through the lesson sub-head thoroughly you will get your answers in it.

(c) This Check Your Progress will help you to understand the lesson better. Try to write answers for them, but do not submit your answers to the university for assessment. These are for your practice only.

15.7 LET US SUM UP

By going through this lesson it is very true and clear that simulation is a reflection of a real system representing the characteristics and behaviour within a given set of conditions. The most important point in simulation is that simulation technique is considered as a valuable tool because of its wide area of application. The most important approach to solving simulation is the Monte Carlo Simulation which can be solved with the help of probabilistic and deterministic model. The deterministic simulation mode, have the alternatives clearly known in advance where as the probabilistic model is stochastic in nature and all decisions are made under uncertainty.

15.8 LESSON-END ACTIVITIES

1. Apply the Monte Carlo Simulation technique weather in forecasting.
2. In the corporate the top Bosses use to take major decisions apply the Simulation techniques in designing and performing organisations take an industry like Reliance, Tata, Infosys to support your answer.

15.9 KEYWORDS

Simulation : A management science analysis that brings into play a construction and mathematical model that represents a real-world situation.

Random number : A number whose digits are selected completely at random.

Flow chart : A graphical means of representing the logic of a simulation model.

15.10 QUESTIONS FOR DISCUSSION

1. **Write True or False against each statement**
 - (a) Simulations models are built for management problems and require management input.
 - (b) All simulation models are very expensive.
 - (c) Simulation is best suited to analyse complex & large practical problem
 - (d) Simulation-generate optimal solution.
 - (e) Simulation model can not be very expensive.
2. **Fill in the blank**
 - (a) Simulation is one of the most widely used _____ analysis book.
 - (b) Simulation allow, for the _____ of real world complications.
 - (c) System _____ in similar to business gaming.
 - (d) Monte Carlo method used _____ number.
 - (e) Simulation experiments generate large amount of _____ and information.
3. **Briefly comment on the following**
 - (a) The problem tackled by simulation may range from very simple to extremely complex.
 - (b) Simulations allows us to study the interactive effect of individual components or variables in order to determine which one is important.

- (c) Simulation is the valuable technique for analysing various maintenance policies before actually implementing them,
- (d) Simulations technique is considered as a valuable tool because of its wide area of application.
- (e) Simulation is nothing more or less than the technique of performing sampling experiment on the model of the system.

Simulation

15.11 TERMINAL QUESTIONS

1. What is simulation? Give a few areas of its application.
2. With the help of a flow chart, briefly explain the simulation process.
3. What are the advantages and limitations of simulation?
4. What is Monte Carlo simulation?
5. Explain the procedure of simulation using random numbers.
6. Explain how simulation is useful in solving queuing and inventory problems.

Exercise Problem

1. A sweet stall observed that the demand for item Mysorpa per week in one kilogram pack is as follows:

Demand / week (per kilo pack)	5	10	15	20	25	30
Frequency	4	22	16	42	10	6

Generate the demand for the next 10 weeks, and also find the average demand.

2. At a service station, cars arrive for water-wash daily. The probability of number of cars that arrive are given in the table below. Simulate the number of cars that will arrive for the next 10 days. Use the following random numbers: 87, 01, 74, 11, 46, 82, 59, 94, 25 and 34.

Cars arrival per day	5	6	7	8	9	10
Probability	0.2	0.15	0.3	0.25	0.05	0.05

3. A private bank has installed an ATM in the city bazaar area. It was found that the time between an arrival and completion of transaction varies from one minute to seven minutes. The arrival and service distribution times are given below. Simulate the ATM operations for the next 30 arrivals.

Time (minutes)	Probability	
	Arrival	Service
1-2	0.10	0.05
2-3	0.15	0.15
3-4	0.30	0.30
4-5	0.25	0.20
5-6	0.10	0.15
6-7	0.10	0.15

Use Monte-Carlo simulation technique and determine:

- a. Waiting time of the customers.
- b. Idle time of the ATM.

4. The materials manager of a firm wishes to determine the expected mean demand for a particular item in stock during the re-order lead time. This information is needed to determine how far in advance to re-order, before the stock level is reduced to zero. However, both the lead time, and the demand per day for the item are random variables, described by the probability distribution.

Lead time (days)	Probability	Demand / day (units)	Probability
1	0.45	1	0.15
2	0.30	2	0.25
3	0.25	3	0.40
4		4	0.20

Manually simulate the problem for 30 re-orders, to estimate the demand during lead time.

5. A company has the capacity to produce around 300 bikes per day. Daily production varies from 295 to 304 depending upon getting the clearance from the final inspection department. The probability distribution of bikes passed through final inspection per day is given below:

Production per day	Probability
295	0.03
296	0.04
297	0.10
298	0.20
299	0.25
300	0.15
301	0.09
302	0.07
303	0.05
304	0.02

The finished bikes are transported in a long trailer lorry sufficient to accommodate 300 mopeds. Simulate the process for 10 days and find:

- a. The average number of bikes waiting in the factory yard.
 - b. The average empty space in the lorry.
6. In a single pump petrol station, it was observed that the inter-arrival times and service times are as given in the table. Using the random numbers given, simulate the queue behaviour for a period of 30 minutes and estimate the probability of the pump being idle and the mean time spent by a customer waiting to fill petrol.

Inter-arrival time		Service time	
Minutes	Probability	Minutes	Probability
1	0.10	2	0.10
3	0.17	4	0.23
5	0.35	6	0.35
7	0.23	8	0.22
9	0.15	10	0.10

Use the following random numbers: 93, 14, 72, 10, 21, 81, 87, 90, 38, 10, 29, 17, 11, 68, 10, 51, 40, 30, 52 & 71.

7. A one-man TV service station receives TV sets for repair. TV sets are repaired on a ‘first come, first served’ basis. The observations of the study made over a 100 day period are given below.

Simulation

No. of TV sets requiring service	Service	
	Frequency of request	
1	15	
2	15	
3	20	
4	25	
5	25	
No. of TV sets serviced	Servicing done	
	Frequency of service	
1	10	
2	30	
3	20	
4	15	
5	25	

Simulate a 10 day period of arrival and service pattern.

8. ABC company stocks certain products. The following data is available:

- a. No. of Units: 0 1 2 3
 Probability: 0.1 0.2 0.4 0.3
- b. The variation of lead time has the following distribution
 Lead time (weeks): 1 2 3
 Probabilities: 0.30 0.40 0.30

The company wants to know (a) how much to order? and (b) when to order ? Assume that the inventory in hand at the start of the experiment is 20 units and 15 units are ordered closed as soon as inventory level falls to 10 units. No back orders are allowed. Simulate the situation for 25 weeks.

9. A box contains 100 balls of which 20 percent are white, 30 percent are black and the remaining are red. Simulate the process for drawing balls at random from the box, identify and note the colour and then replace. Use the following 10 random numbers to simulate: 52, 60, 02, 3379, 79, 30, 36, 58 and 43.
10. Rahul, the captain of the cricket team, has the following observations on the number of runs scored against type of ball. The bowling probability of a bowler for the type of balls bowled are given below.

Type of bowling	Probability of hitting a boundary
Over pitched	0.1
Short-Pitched	0.3
Outside off stump	0.2
Outside leg stump	0.15
Bouncer	0.20
Attempted Yorker	0.05

The number of runs scored off each type of ball is shown in the table given below:

Type of bowling	Probability of hitting a boundary
Over pitched	1
Short-Pitched	4
Outside off stump	3
Out side leg stump	2
Bouncer	2
Attempted Yorker	0

Simulate the game for 3 overs (6 balls per over) and calculate the batting average of Rahul.

15.12 MODEL ANSWERS TO QUESTIONS FOR DISCUSSION

1. (a) True (b) False (c) True (d) False (e) False
2. (a) Quantitative (b) Inclusion (c) Simulation (d) Random (e) Data

15.13 SUGGESTED READINGS

Ernshoff, J.R. & Sisson, R.L. *Computer Simulations Models*, New York Macmillan Company.

Gordon G., *System Simulation*, Englewood cliffs N.J. Prentice Hall.

Chung, K.H. "Computer Simulation of Queuing System" Production & Inventory Management Vol. 10.

Shannon, R. I. *Systems Simulation. The art & Science*. Englewood Cliffs, N.J. Prentice Hall.