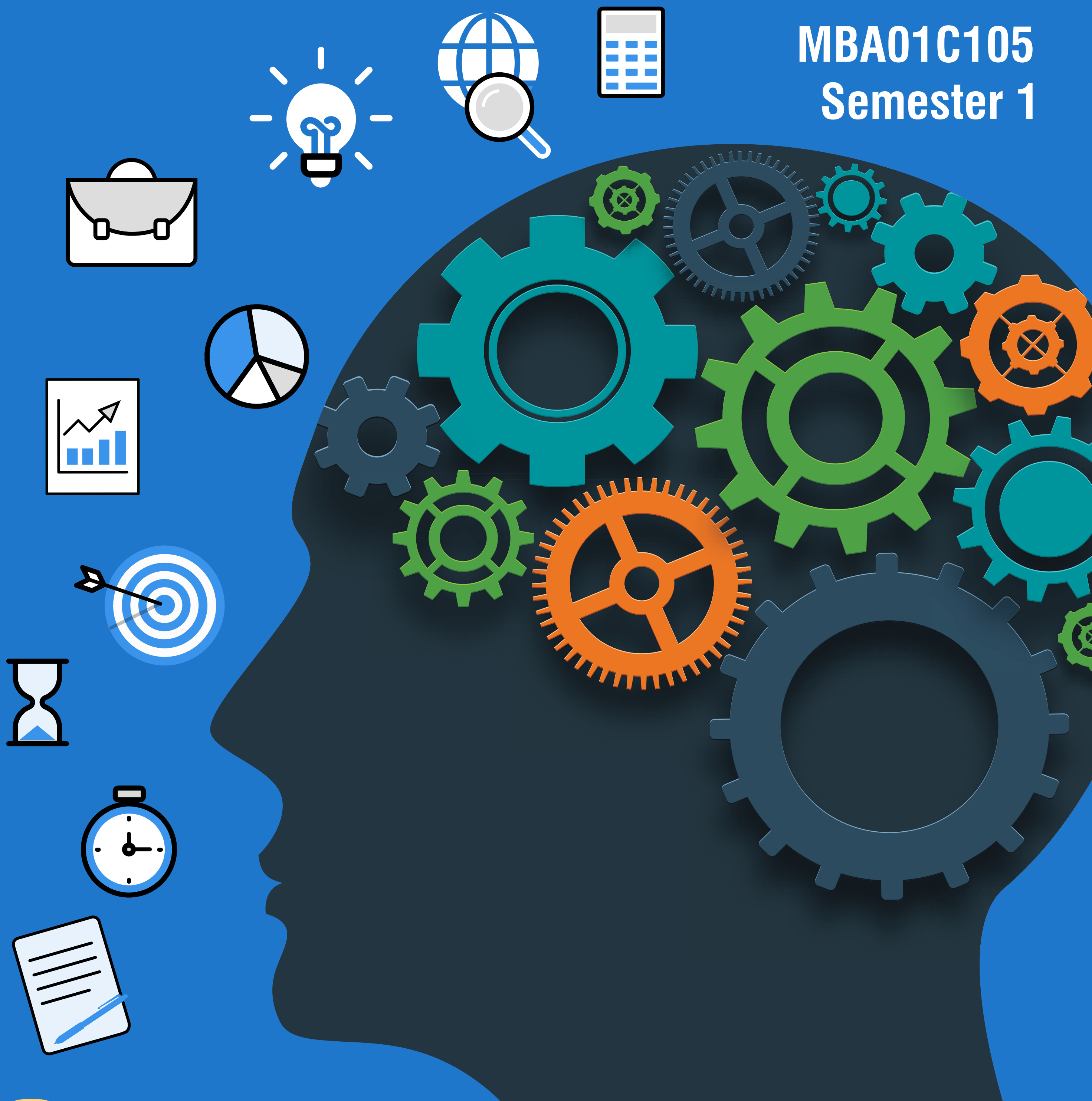
BAOU
Education for All

DR. BABASAHEB AMBEDKAR
OPEN UNIVERSITY

MBA01C105
Semester 1



QUANTITATIVE MANAGEMENT

Message for the Students

Dr. Babasaheb Ambedkar Open (University is the only state Open University, established by the Government of Gujarat by the Act No. 14 of 1994 passed by the Gujarat State Legislature; in the memory of the creator of Indian Constitution and Bharat Ratna Dr. Babasaheb Ambedkar. We Stand at the seventh position in terms of establishment of the Open Universities in the country. The University provides as many as 54 courses including various Certificate, Diploma, UG, PG as well as Doctoral to strengthen Higher Education across the state.

On the occasion of the birth anniversary of Babasaheb Ambedkar, the Gujarat government secured a quiet place with the latest convenience for University, and created a building with all the modern amenities named 'Jyotirmay' Parisar. The Board of Management of the University has greatly contributed to the making of the University and will continue to this by all the means.

Education is the perceived capital investment. Education can contribute more to improving the quality of the people. Here I remember the educational philosophy laid down by Shri Swami Vivekananda:

> ***"We want the education by which the character is formed, strength of mind is Increased, the intellect is expand and by which one can stand on one's own feet".***

In order to provide students with qualitative, skill and life oriented education at their threshold. Dr. Babaasaheb Ambedkar Open University is dedicated to this very manifestation of education. The university is incessantly working to provide higher education to the wider mass across the state of Gujarat and prepare them to face day to day challenges and lead their lives with all the capacity for the upliftment of the society in general and the nation in particular.

The university following the core motto 'स्वाध्यायः परमम् तपः' does believe in offering enriched curriculum to the student. The university has come up with lucid material for the better understanding of the students in their concerned subject. With this, the university has widened scope for those students who are not able to continue with their education in regular/conventional mode. In every subject a dedicated term for Self Learning Material comprising of Programme advisory committee members, content writers and content and language reviewers has been formed to cater the needs of the students.

Matching with the pace of the digital world, the university has its own digital platform Omkar-e to provide education through ICT. Very soon, the University going to offer new online Certificate and Diploma programme on various subjects like Yoga, Naturopathy, and Indian Classical Dance etc. would be available as elective also.

With all these efforts, Dr. Babasaheb Ambedkar Open University is in the process of being core centre of Knowledge and Education and we invite you to join hands to this pious *Yajna* and bring the dreams of Dr. Babasaheb Ambedkar of Harmonious Society come true.

Prof. Ami Upadhyay
Vice Chancellor,
Dr. Babasaheb Ambedkar Open University,
Ahmedabad.

MBA

SEMESTER-1

QUANTITATIVE MANAGEMENT

BLOCK: 1

Authors' Name:	Dr. Umesh Raval Dr. Paresh Andhariya
Review (Subject):	Prof. (Dr.) Manoj Shah Dr. Ravi Vaidya Dr. Maulik Desai
Review (Language):	Dr. Jainee Shah
Editor's Name:	Prof. (Dr.) Manoj Shah, Professor and Director, School of Commerce and Management, Dr. Babasaheb Ambedkar Open University, Ahmedabad.
Co-Editor's Name:	Dr. Dhaval Pandya Assistant Professor, School of Commerce and Management, Dr. Babasaheb Ambedkar Open University, Ahmedabad.

Publisher's Name:	Registrar, Dr. Babasaheb Ambedkar Open University, 'JyotirrmayParisar',Opp. Shri Balaji Temple,Chharodi, Ahmedabad, 382481, Gujarat, India.
Edition:	2022 (First Edition)
ISBN:	978-93-5598-343-5

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means without permission in writing from Dr. Babasaheb Ambedkar Open University, Ahmedabad.

Dr. Babasaheb Ambedkar Open University
(Established by Government of Gujarat)

QUANTITATIVE MANAGEMENT
SEMESTER-1

BLOCK
1

BLOCK
2

UNIT 01	QUANTITATIVE TECHNIQUE – INTRODUCTION

1.1 INTRODUCTION TO STATISTICS

Now-a-days the word 'statistics' has become a household word, although different people comprehend it in different senses. The modern educated person has to be a person of statistics, broadly understanding its meaning and applying it to his/her life in different ways. For example, every day we come across different types of quantitative information in both print as well as electronic media on topics like population, exchange rate fluctuations, inflation rate, day and night temperatures (being below or above normal; lowest or highest in the century or in the last thirty years or so), etc. In order to improve our understanding of the world around us, it is necessary to:

a) Measure what is being said,
b) Express it numerically, i.e., in numbers/quantities like weights in so many kilograms, eggs in somany dozens, etc.
c) Utilize quantitative information or expression to draw conclusions and suggest policy measures.

Needless to say that if we cannot measure and express, in terms of numbers what is being said, then our knowledge will remain insufficient and far from being satisfactory. Statistics thus involves some sort of numerical information called "numerical data" or simply "data". For example, one may give a statement that he/she has studied statistics (that is quantitative information) on absenteeism the educated and the uneducated workers in Indian industries and found that incidence of absenteeism is more among the latter. He/she is referring to the numerical figures or numerical information technically called data.

Other examples of data are:

a) India is suffering from population explosion, annual growth of population being around 1%.
b) Foreign exchange reserve of the country has been the highest so far since independence and stoodat $ 110 billion.
c) Students of 12^{th}-A have shown a better result than those of 12^{th}-B because the average marks ofthe former are 38% more than the average marks of the latter.

Many more such examples can be found and the students are expected to go through this exercise on theirown.

1.2 IMPORTANCE OF STATISTICS IN MODERN BUSINESS STATISTICS

A person managing a production unit, farm, factory or domestic kitchen has to coordinate people, machines and money against several constraints, like those of time, cost and space, in order to achieve the organization's objectives in an efficient and effective manner. The manager has to analyze the situation on a continuous basis; determine the objectives; identify the best option from the set of available alternatives; and implement, coordinate, evaluate and control the situation continuously to achieve these objectives.

Organizations of today have become increasingly complex, and hence managerial decision making has become even more complex. As a result, management is becoming more of a science than an art.

As the complexity of organizations and the business environment has made the process of decision making difficult, managers cannot take decisions on the basis of subjective factors like their experience, observation or evaluation anymore. Decisions need to be based on thorough analysis of data that reveals relationships, indicates trends and shows rates of change in the relevant variables.

Quantitative methods provide ways to collect, present, analyze and interpret the available data meaningfully. They are the powerful tools through which managers can accomplish their predetermined objectives, like profit maximization, cost minimization or efficient and effective use of production capacities. The study of quantitative methods has a wide range of applications, especially in business.

The information needed by the decision maker or owner to make effective decisions was much less extensive. Thus, he made decisions based on his past experience and intuition only. The reasons are

- The marketing of the product was not a problem because customers were personally known to the business owner.
- Test marketing of the product was not essential because the owner used to know the choice and need of the customers just by interaction.
- The owner used to work with his workers at the shop floor. He knew all of them personally.

- Progress on the work was being made daily at the work center. Thus, production records were not required.

❖ **Quantitative methods are used in decision making for the following reasons:**

- Complexity of today's managerial activities, which involve constant analysis in setting objectives, seeking alternatives and implementing, coordinating, controlling and evaluating the result.
- Availability of different tools for quantitative analysis of complex problems

❖ **Business Statistics helps a business to:**

- Deal with uncertainties by forecasting seasonal, cyclic and general economic fluctuations
- Helps in Sound Decision making by providing accurate estimates about costs, demand, prices, sales etc.
- Helps in business planning on the basis of sound predictions and assumptions
- Helps in measuring variations in performance of products, employees, business units, etc.
- It allows comparison of two or more products, business units, sales teams, etc.
- Helps in identifying relationship between various variables and their effect on each other like effect of advertisement on sales
- Helps in validating generalizations and theoretical concepts formulated by managers

1.3 MEANING AND DEFINITION OF STATISTICS

1.3.1 Meaning of Statistics

"Statistics", that a word is often used, has been derived from the Latin word "Status" that means a group of numbers or figures; those represent some information of our human interest.

We find statistics in everyday life, such as in books or other information papers or TV or newspapers.

Although, in the beginning it was used by Kings only for collecting information about states and other information which was needed about their people, their number, revenue of the state etc. This was known as the science of the state because it was used only by the Kings. So it got its development as "Kings" subject or "Science of Kings" or we may call it as "Political Arithmetic's". It was for the first time, perhaps in Egypt to conduct census of population in 3050 B.C. because the king needed money to erect pyramids. But in India, it is thought, that, it started dating back to Chandra Gupta Maurya's kingdom under Chanakya to collect the data of births and deaths. TM has also been stated in Chanakya's Arthshastra.

But now-a-days due to its pervading nature, its scope has increased and widened. It is now used in almost in all the fields of human knowledge and skills like Business, Commerce, Economics, Social Sciences, Politics, Planning, Medicine and other sciences, Physical as well as Natural.

1.3.2 History of Statistics

The word Statistics is the modern form of the word Statistik which in turn has been derived from the Italian word "Statista" meaning "statesman". Professor GottFried Adren Wall used it in the 18th century. It was Dr. E.A.W. Zimmerman who introduced the word statistics into England.

Early government records show statistical information on some aspects of epidemics and so on. Perhaps it was because of this that Statistics was called the population, land records, military strength of different wings and mortality during science of kings. But as the humanity developed, the usage as well as understanding of Statistics increased and now it is difficult to imagine a field of knowledge which can do without statistics. In fact, it has become an important tool of analysis.

1.3.3 Definition of Statistics

The term "Statistics" has been defined in two senses, i.e. in Singular and in Plural sense.

"Statistics has two meanings, as in plural sense and in singular sense". **- Oxford Dictionary**

In plural sense, it means a systematic collection of numerical facts and in singular sense; it is the science of collecting, classifying and using statistics.

A. In the Plural Sense:

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

- "The classified facts respecting the condition of the people in a state-especially those facts which can be stated in numbers or in tables of numbers or in any tabular or classified arrangement." **- Webster**

Definitions given above give a narrow meaning to the statistics as they do not indicate its various aspects as are witnessed in its practical applications. From this point of view the definition given by Prof. Horace Sacrist appears to be the most comprehensive and meaningful:

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

B. In the Singular Sense:

- "Statistics refers to the body of technique or methodology, which has been developed for the collection, presentation and analysis of quantitative data and for the use of such data in decision making." **- Ncttor and Washerman**

- "Statistics may rightly be called the science of averages." **- Bowleg**

- "Statistics may be defined as the collection, presentation, analysis, and interpretation of numerical data." **- Croxton and Cowden**

C. Some Modern Definitions:

From the above two senses of statistics, modem definitions have emerged as given below:

- "Statistics is a body of methods for making wise decisions on the face of uncertainty." **- Wallis and Roberts**
- "Statistics is a body of methods for obtaining and analyzing numerical data in order to make better decisions in an uncertain world." **- Edward N. Dubois**

So, from above definitions we find that science of statistics also includes the methods of collecting, organizing, presenting, analyzing and interpreting numerical facts and decisions are taken on their basis. The most proper definition of statistics can be given as following after analyzing the various definitions of statistics.

- "Statistics in the plural sense are numerical statements of facts capable of some meaningful analysis and interpretation, and in singular sense, it relates to the collection, classification, presentation and interpretation of numerical data."

1.4 SCOPE, APPLICATION AND CHARACTERISTICS OF STATISTICS

The use of statistics in modern times is quite adequate and broad. Statistics are numerical statement of facts capable of analysis and interpretation as well as study of the methods used in collection, organization, presentation, analysis and interpretation of numerical data. It is often said that **"Statisticsis what Statisticians do"**. Its scope is unlimited and includes the following:

- Statistics and planning: Statistics in indispensable into planning in the modern age which is term as "the age of planning". Almost all over the world the government is re-storing to planning for economic development.
- Statistics and economics: Statistical data and technique of statistical analysis have to immensely useful involving economical problem. Statistics is helpful in understanding the intensity of the economic problem and deriving the solutions for the same on the basis of the data available. Such as wages price, time series analysis, termed analysis.
- Statistics and Econometrics: Econometrics is one of the most recent fields of study concerning economics. It combines the methods and techniques of Statistics and Mathematics to build models for the analysis of economic problems and then provides solution to all those problems. For example, in Econometrics Linear Regression models are formulated to analyze the effect of various determinants of demand on demand for a product.
- Statistics and Business: Statistics is an irresponsible tool of production control. Business executive are relying more and more on statistical technique for studying the much and desire of valued customers.
- Statistics and industry: In industry is widely used inequality control. In production, engineering to find out whether the product is confirming to the specification or not. For that Statistical Quality

Control is a branch of Statistics that deals with "Quality Control" of manufactured goods. With the help of probability theory and sampling technique, control charts and inspection plans are formulated. Statistical tools, such as inspection plan, control chart, etc.

- Statistics and Mathematics: Statistics are intimately related recent advancement in statistical techniques is the outcome of wide application of Mathematics.
- Statistics and Modern science: In medical science the statistical tools for collection and incidence of diseases and result of application various drugs and Medicines are of great importance.

We saw some of the area in which statistics is used. But it is not end of scope of statistics. Statistics is widely used in many other branches like Natural Science, Social Science, Space Science, Agricultural Science, Physical Science, Research and Planning, Banking and Insurance, Commerce, etc.

1.4.1 Applications of Statistics:

Marketing: As per Philip Kotler and Gary Armstrong marketing "identifies customer needs and wants, determine which target markets the organizations can serve best, and designs appropriate products, services and Programs to serve these markets". Marketing is all about creating and growing customers profitably. Statistics is used in almost every aspect of creating and growing customers profitably. Statistics is extensively used in making decisions regarding how to sell products to customers. Also, intelligent use of statistics helps managers to design marketing campaigns targeted at the potential customers. Marketing research is the systematic and objective gathering, recording and analysis of data about aspects related to marketing. Web analytics is about the tracking of online behavior of potential customers and studying the behavior of browsers to various websites. Use of Statistics is indispensable in forecasting sales, market share and demand for various types of Industrial products. Factor analysis, conjoint analysis and multidimensional scaling are invaluable tools which are based on statistical concepts, for designing of products and services based on customer response.

Finance: Uncertainty is the hallmark of the financial world. All financial decisions are based on "Expectation" that is best analyzed with the help of the theory of probability and statistical techniques. Probability and statistics are used extensively in designing of new insurance policies and in fixing of premiums for insurance policies. Statistical tools and technique are used for analyzing risk and quantifying risk, also used in valuation of derivative instruments, comparing return on investment in two or more instruments or companies. Beta of a stock or equity is a statistical tool for comparing volatility, and is highly useful for selection of portfolio of stocks. The most sophisticated traders in today's stock markets are those who trade in "derivatives" i.e. financial instruments whose underlying price depends on the price of some other asset.

Economics: Statistical data and methods render valuable assistance in the proper understanding of the economic problem and the formulation of economic policies. Most economic phenomena and indicators can be quantified and dealt with statistically sound logic. In fact, Statistics got so much integrated with Economics that it led to development of a new subject called Econometrics which basically deals with economics issues involving use of Statistics.

Operations: The field of operations is about transforming various resources into product and services in the place, quantity, cost, quality and time as required by the customers. Statistics playsa very useful role at the input stage through sampling inspection and inventory management, in the process stage through statistical quality control and six sigma method, and in the output stage through sampling inspection. The term Six Sigma quality refers to situation where there are only.

Human Resource Management or Development: Human Resource departments are inter alia entrusted with the responsibility of evaluating the performance, developing rating systems, evolving compensatory reward and training system, etc. All these functions involve designing forms, collecting, storing, retrieval and analysis of a mass of data. All these functions can be performed efficiently and effectively with the help of statistics.

Information Systems: Information Technology (IT) and statistics both have similar systematic approach in problem solving. IT uses Statistics in various areas like, optimization of server time,assessing performance of a program by finding time taken as well as resources used by the Program. It is also used in testing of the software.

Data Mining: Data Mining is used in almost all fields of business.

In Marketing, Data mining can be used for market analysis and management, target marketing, CRM, market basket analysis, cross selling, market segmentation, customer profiling and managing web based marketing, etc.

In Risk analysis and management, it is used for forecasting, customer retention, quality control, competitive analysis and detection of unusual patterns.

In Finance, it is used in corporate planning and risk evaluation, financial planning and asset evaluation, cash flow analysis and prediction, contingent claim analysis to evaluate assets, cross sectional and time series analysis, customer credit rating, detecting of money laundering and other financial crimes.

In Operations, it is used for resource planning, for summarizing and comparing the resources andspending.

In Retail industry, it is used to identify customer behaviors, patterns and trends as also fordesigning more effective goods transportation and distribution policies, etc.

1.4.2. Characteristics of Statistics:

As per we seen definitions earlier, clearly points out certain **characteristics** which numerical data must possess in order that they may be called statistics. These are as follows:

1. **Statistics are aggregates of facts:** Single and isolated figures are not statistics because they

cannot be compared and no meaningful conclusion can be drawn from it. It is the only aggregate of facts capable of offering some meaningful conclusion that constitute statistics. (All statistics are expressed in numbers but all numbers are not statistics)

2. **Statistics must be numerically expressed:** Statistical methods are applicable only to those data which can be numerically expressed. Qualitative expressions like honesty, intelligence, sincere are not statistics unless they can be numerically expressed.

3. **Statistics should be capable of being related to each other:** Statistical data should be capable of comparison and connected to each other. If there is no apparent relationship between the data they cannot be called statistics.

4. **Statistics should be collected in a systematic manner:** For collecting statistical data a suitable plan should be prepared and work should be done accordingly.

5. **Statistics should be collected for a definite purpose:** The purpose of collecting data must be decided in advance. The purpose should be specific and well defined.

6. **Statistics are affected to a marked extent by a large number of causes:** Facts and figures are affected to a marked extent by the combined influence of a number of forces.

7. **Reasonable standard of accuracy should be maintained in collection of statistics:** Statistics deals with large number of data. Instead of counting each and every item, Statisticians take a sample and apply the result thus obtained from sample to the whole group. The degree of accuracy of sample largely depends upon the nature and object of the enquiry. If reasonable standard of accuracy is not maintained, numbers may give misleading result.

1.5 FUNCTION OF STATISTICS

The functions of statistics are as follows:

1. **It presents fact in a definite form:** Numerical expressions are convincing and, therefore, one of the most important functions of statistics is to present statement in a precise and definite form.
2. **It simplifies mass of figures:** The data presented in the form of table, graph or diagram, average or coefficients are simple to understand.
3. **It facilitates comparison:** Once the data are simplified they can be compared with other similar data. Without such comparison the figures would have been useless.
4. **It helps in prediction:** Plans and policies of organizations are invariably formulated in advance at the time of their implementation. Knowledge of future trends is very useful in framing suitable policies and plans.
5. **It helps in formulating and testing hypothesis:** Statistical methods like z-test, t-test, X^2(Chi-Square)-test are extremely helpful in formulating and testing hypothesis and to develop new theories.
6. **It helps in the formulation of suitable policies:** Statistics provide the basic material for framing

suitable policies. It helps in estimating export, import or production programmes in the light of changes that may occur.

7. **Statistics indicates trend behavior:** Statistical techniques such as Correlation, Regression, Time series analysis etc. are useful in forecasting future events.

1.6 LIMITATIONS OF STATISTICS

Limitations of statistics are as follows:

1. **Statistics deals only with quantitative characteristics:** Statistics are numerical statements of facts. Data which cannot be expressed in numbers are incapable of statistical analysis. Qualitative characteristics like honesty, efficiency, intelligence, etc. cannot be studied directly.
2. **Statistics deals with aggregates not with individuals:** Since statistics deals with aggregates of facts, the study of individual measurements lies outside the scope of statistics.
3. **Statistical laws are not perfectly accurate:** Statistics deals with such characteristics which are affected by multiplicity of causes and it is not possible to study the effect of these factors. Due to this limitation, the results obtained are not perfectly accurate but only an approximation.
4. **Statistical results are only an average:** Statistical results reveal only the average behavior. The Conclusions obtained statistically are not universally true but they are true only under certain conditions.
5. **Statistics is only one of the methods of studying a problem:** Statistical tools do not provide the best solution under all circumstances.
6. **Statistics can be misused:** The greatest limitation of statistics is that they are liable to be misused. The data placed to an inexperienced person may reveal wrong results. Only persons having fundamental knowledge of statistical methods can handle the data properly.

1.7 STATISTICAL SOFTWARE

To perform complex statistical analysis; organization, interpretation and presentation of big data set without wasting time we use statistical software. There are many statistical software are available now-a-days in the market, even online statistical software are also available. We shall see some of popular statistical software.

1. **Excel:** Excel is spreadsheet applications that can be used to cover some basic statistical analyses, especially when you utilize the most recent statistics add in. However complex multivariate analyses such as Factor, Multivariate Regression, Cluster, Multi Dimensional Scaling, Structural Equation Modelling, et. al. analyses will require a dedicated stats package such as SPSS or SAS. Excel can be used in principle to do some simple linear regression, correlations, some hypothesis testing, and has some helpful charting capabilities, but all up you are better off with a stats package.

2. **SPSS:** SPSS Statistics is a software package used for interactive or batched, statistical analysis. The software name originally stood for **Statistical Package for the Social Sciences (SPSS)**, reflecting the original market and then later changed to **Statistical Product and Service Solutions (SPSS)**. SPSS is a widely used program for statistical analysis in social science, market researchers, health researchers, survey companies, government, education researchers, marketing organizations, data miners and others. The original SPSS manual (Nie, Bent & Hull, 1970) has been described as one of **"sociology's most influential books"** for allowing ordinary researchers to do their own statistical analysis. In addition to statistical

analysis, data management (case selection, file reshaping, creating derived data) and data documentation (a metadata dictionary is stored in the data file) are features of the base software.

3. **R-Language: Business Analytics with R** or commonly known as **"R Programming Language"** is an open-source programming language and a software environment designed by and for statisticians. It is basically used for **statistical computations** and **high-end graphics**. Thus, it is a popular language among mathematicians, statisticians, data miners, and also scientists to do data analysis. **R is a GNU project**, and is freely available under the GNU (General Public License), and R comes with pre-compiled binary versions for several operating systems ranging from Unix and similar systems (FreeBSD, Linux), Windows and also MacOS.

4. **Stata:** Stata is a statistical analysis solution designed to help businesses streamline data analysis, manipulation, visualization and management. It allows businesses to create, merge, sort and merge multiple datasets, import/export data from **Excel** and **.CSV** formats and automatically adjust the platform's memory usage according to data requirements. Key features of Stata include version control, data management, built-in spreadsheet, SEM diagram building and variable management. It comes with an integrated graph editor, which lets businesses create custom graphs using titles, lines, arrows, notes and texts. Additionally, it includes an automated reporting functionality, allowing users to produce analysis documents and reports in Word, PDF, HTML, SVG, Excel and PNG formats.

5. **Python:** Business Intelligence aims to help businesses draw insights from their past data to make better decisions for the future. BI tools are often built on Python to access, classify, and process data to derive useful outcomes. The Business Intelligence dashboards mostly built on Python creates a visual interface for the business to define and track metrics and KPIs. Unlike in the case of excel spreadsheets, you do not have to hover across various documents to observe their business data; Business Intelligence dashboards give a comprehensive view of your overall business. Python facilitated predictive analysis which is the branch of data science and machine learning, where the data of past events is analyzed to predict future outcomes. Prescriptive Analytics is a subsidiary of Business Intelligence. It is the science of anticipating what outcome will occur when it will occur and why it will occur and what possibly can your business do with this information. It uses your business data and applies it to the decision-making process to help you make data-driven insightful decisions.

6. **SAS: SAS** is a statistical software suite developed by SAS Institute for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation and predictive analytics. SAS is a software suite that can mine, alter, manage and retrieve data from a variety of sources and perform statistical analysis on it. SAS provides a graphical point-and-click user interface for non-technical users and more through the SAS language.

There are many software available now a days as per our requirement of study. We have seen some of the popular softwares.

❖ Key Words:

Statistics: In plural sense, it means a set of numerical figures commonly known as statistical data.

Statistics: In singular sense, it means scientific methods for collection, presentation, analysis and interpretation of data.

Statistic: It is measure, like arithmetic mean, median, geometric mean, standard deviation, etc., calculated from sample. It is also termed as estimator in the theory of estimation.

Parameter: It is a measure like arithmetic mean, median, geometric mean, standard deviation, etc., calculated by using all values of population.

Quantitative Data: There are information on measurable characteristics. Such data are available in the form of numerical figures.

Qualitative data: These are information on non-measurable characteristics like honesty, beauty, color, caste, etc.

Census: A method of investigation in which information is collected from all units of the population.

Sampling: A method of investigation in which information is collected from sampled units only.

❖ CHECK YOUR PROGRESS

- **Answer the following Multiple Choice Questions**

Que. 1 Who used the German word "Statistik" for the first time?

(a) John Graunt (b) Gottfried Adren Wall (c) Willian Patty (d) Gauss

Que. 2 Which of the following is an example of Quantitative Data?

(a) Income (b) Gender (c) Religion (d) Citizenship

Que. 3 Which one of the following is not a function of statistics?

(a) To simplify complexities (b) To forecast the future

(c) To compare data with respect to time and date (d) To pass a bill

Que. 4 Statistical methods are

(a) Collection of data (b) Analysis of data (c) Classification of data (d) All of these

Que. 5 In_________sense, statistics refers to a set of methods and techniques used for collection, tabulation, analysis and interpretation of statistical data.

(a) Normal (b) Singular (d) Plural (d) Varied

Que. 6 Statistical results are

(a) Absolutely correct (b) Universally Correct

(c) True on an average (d) Not correct

Que. 7 Statistics can

(a) Prove anything (b) Disprove anything

(c) Neither prove nor disprove anything: but is a tool (d) solve everything

- **Answer the following Question in one Sentence**

Que. 1 Define Quantitative data.
Que. 2 Define Qualitative Data.
Que. 3 Define Statistics (In Singular Sense).
Que. 4 Define Statistics (In Plural Sense).
Que. 5 Give name of Statistical Software. (any four)
Que. 6 Write any one function of Statistics.

- **Answer the following Question in detail**

Que. 1 Give meaning and history of Statistics.
Que. 2 What is Statistics? Give the definition of Statistics.
Que. 3 Write Characteristics of Statistics.
Que. 4 State the Limitations of Statistics.
Que. 5 Write application of Statistics in various field.
Que. 6 List at least three applications of statistics in each functional area.
Que. 7 Comment on the following statements:

a. "Statistics are numerical statement of facts, but all facts numerically stated are not statistics".
b. "Statistics is the science of averages"

- **For Further Reading Some Reference Books:**

- Elhance, D.N. and V. Elhance, 1988, Fundamental of Statistics, Kitab Mahal, Allahabad.
- Nagar, A.L. and R.K. Dass, 1983, Basic Statistics, Oxford University Press, Delhi
- Mansfield, E., 1991, Statistics for Business and Economics: Method and Applications, W.W. Norton and Co.
- Yule, G.U. and M.G Kendall, 1991, An Introduction to the Theory of Statistics, University Books, Delhi.
- Umeshkumar Dubey, D P Kothari, G K Awari, Quantitative Techniques in Business, Management and Finance, A Case Study approach, Taylor and Francis Group.
- Robert V. Hogg, Joseph W. McKean, Allen T. Craig, Introduction to Mathematical Statistics, Eight Editions.
- Timothy C. Haas, Introduction to Probability and Statistics for Ecosystem Managers, WILEY

UNIT 02

MEASURES OF CENTRAL TENDENCY

2.1 INTRODUCTION

For proper understanding of quantitative data, it should be classified and converted into a frequency distribution, the number of times or frequency with which a particular data occurs in the given mass of data. This type of condensation of data reduces its bulk and gives a clear picture of its structure.

One of the most important objectives of statistical analysis is to get one single value that describes the characteristic of the entire mass of unwieldy data. Such a value is called the central value or an 'average' or the expected value of the variable. The word average is very commonly used in day-to-day conversation. For example, we often talk of average boy in a class, average height or life of an Indian, average income, etc. When we say 'he is an average student' what it means is that he is neither very good nor very bad, just a mediocre type of student. However, in statistics the term average has a different meaning.

2.1.1 AVERAGE DEFINED

The word 'average' has been defined differently by various authors. Some important definitions are givenbelow:

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

"A measure of Central tendency is a typical value around which other figures congregate." **– Simpson and Kafka**

"An average is a single value selected from a group of values to represent them in some way – a value which is supposed to stand for whole group, of which it is a part, as typical of all the values in the group."
– A.E. Waugh

"An average is a typical value in the sense that it is sometimes employed to represent all the individual values in a series or of a variable." **– Ya-Lun-Chou**

"The average is sometimes described as a number which is typical of the whole group."
– Leabo

"An average value is a single value within the range of the data that is used to represent all of the values in the series. Since an average is somewhere within the range of the data, it is also called a measure of central value." **– Croxton & Cowden**

It is clear from the above definitions that an average is a single value that represents a group of values. Such a value is of great significance because it depicts the characteristic of the whole group. In other words, average is a statistical measure representing a group of individual values in simple and comprehensive manner. Since an average represents the entire data, its value lies somewhere in between the two extremes, i.e., the largest and the smallest items. For this reason an average is frequently referred to as a measure of central tendency.

2.1.2 OBJECTIVES OF AVERAGING:

(i) **Representative of the group:** An average represents all the features of a group; hence the resultsabout the whole group can be deduced from it.

(ii) **Brief Description:** An average gives us simple and brief description of the main features of thewhole data.

(iii) **Helpful in Comparison:** The measures of central tendency or averages reduce the data to a single value which is highly useful for making comparative studies. For example, comparing theper capita income of two countries, we can conclude that which country is richer.

(iv) **Helpful in Formulation of Policies:** Averages help to develop a business in case of a firm orhelp the economy of a country to develop.

(v) **Base of other Statistical Analysis:** Other statistical devices such as mean deviation, co-efficient of variation, co-relation, analysis of time series and index numbers are also based on the averages.

2.2 IMPORTANCE OF AVERAGE

It is also called measure of central tendency i.e. a measure of the central value in the population.It is representative of the entire data. If X is the average of a dataset, then the numbers to its left and right balance each other. It is easily affected by outliers. It is a term used for discrete random variables whereas for continuous random variables, the term used is Expected value. They are widely used in sociology, economics, meteorology, etc. This concept is used in our day to day lives as well here we want to understand how much money we spend per week or fuel consumed per mile, etc.

There are several examples which highlight important of average in our day to day lives.

1. The mileage of vehicles or cars is an average value. This shows the performance of a car i.e. how much fuel is used for the distance traveled. If the car covers a larger distance on lower consumption, it is a good car.
2. In cricket, batting average is the number of runs the batsman scores to the number of times he bats. The higher the score per innings show that the player is good as compared to someone who doesn’t score many runs per innings.
3. Calculating averages of marks is a common practice in schools and colleges. This helps decide the rank of students based on the marks across several subjects.

2.3 DIFFERENCE MEASURES: CENTRAL TENDENCY AND DISPERSION

Measures of Central Tendency: A **measure of central tendency** is a number used to represent the center or middle of a set of data values. The **mean**, **median** and **mode** are three commonly used measures of central tendency.

Measures of Dispersion: A **measure of dispersion** is a statistic that tells you how dispersed or spread out, data values are. One simple **measure of dispersion** is the **range**, which is the difference between the greatest and least data values.

2.4 MEASURES OF CENTRAL TENDENCY: MEAN

2.4.1 Definition:

Mean is defined as the value obtained by sum of all observation divided by the total number of observation. Mean is also known as Arithmetic Mean. Mean of variable *x* is denoted by $\bar{x}$.

2.4.2 Advantages and disadvantages of mean:

Advantages:

Mean is the most popular measure of central tendency because of its following advantages.

(1) It is rigidly defined. It has a fixed mathematical formula.
(2) It is easy to understand and calculate.
(3) It is based on all observations.
(4) It is suitable for further algebraic operations.
(5) It is comparatively more stable measure. It means that means of all samples of same size from the same population have comparatively less variation.
(6) All the observations are given equal importance in the calculation of mean.

Disadvantages:

The following disadvantages should be also taken into consideration before using mean as a measure of central tendency.

(1) It is unduly affected by too large or too small values.
(2) It cannot be calculated for data having classes with open ends.
(3) Its exact value cannot be found graphically or by inspection.
(4) If few observations are missing, exact value of mean cannot be found.
(5) Mean is not a good representative value for the data which are not evenly spread around their average.
(6) Using mean as an average is not appropriate if observations have varying importance.

2.4.3 Calculation:

For Ungrouped Data: Suppose $x_1, x_2, \ldots, x_n$ are the *n* observation in the data, then Mean is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

Where, $\sum x_i = x_1 + x_2 + \cdots + x_n = Sum\ of\ observations\ x_1, x_2, \ldots, x_n$ and

$n = number\ of\ observation$

Short cut Method:

If the values of observation are very large, the calculation can be simplified by using assumed mean **A**. A is some constant value, preferably around the center of all the observation. Assumed mean **A** is subtractedfrom observation $x_1, x_2, \ldots, x_n$ and the deviation denoted by $d_1, d_2, \ldots, d_n$. $(d_i = x_i - A; i = 1, 2, \ldots, n)$
then the Mean is

$$\bar{x} = A + \frac{\sum d_i}{n}$$

Where, $\sum \boldsymbol{d_i} = d_1 + d_2 + \cdots + d_n = Sum\ of\ observations\ d_1, d_2, \ldots, d_n$; $\boldsymbol{A} = Assumed\ Mean$

$\boldsymbol{n} = number\ of\ observation$

Illustration 1: The following data gives the monthly sales of 10 salesmen of particular brand Water Purifier. Calculate Mean sales using the data.

Sales (in Pc)	16	19	20	17	23	18	15	21	12	22

Answer:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{16 + 19 + 20 + 17 + 23 + 18 + 15 + 21 + 12 + 22}{10}$$

$$\bar{x} = \frac{183}{10} = 18.3$$

Hence the Mean Sales of Water Purifier is **18.3 pc**.

Illustration 2: The following table gives the monthly income of 10 employees in an office. Calculate Mean (average) income of employee.

Income (Rs.)	1780	1670	1690	1750	1840	1920	1100	1810	1050	1950

Answer:

Here we can see that value of variable is large. So, we subtract Assumed Mean (A = 1800) from all the observation and new variable is denoted by d.

Income (Rs.) $(\boldsymbol{x_i})$	1780	1670	1690	1750	1840	1920	1700	1810	1850	1950
$\boldsymbol{d_i = x_i - 1650}$	**-20**	**-130**	**-110**	**-50**	**40**	**120**	**-100**	**10**	**50**	**150**

$$\bar{x} = A + \frac{\sum d_i}{n}$$

$$= 1800 + \left(\frac{(-20) + (-130) + (-110) + (-50) + 40 + 120 + (-100) + 10 + 50 + 150}{10}\right)$$

$$= 1800 + \left(\frac{-40}{10}\right) = 1800 - 4 = 1796$$

For Grouped Data:

For Discrete frequency distribution:

Suppose x_1, x_2, ... , x_k are the observation in the given data with frequencies f_1, f_2, ... , f_k respectively. Here, $n = total\ number\ of\ observation = \sum f_i = f_1 + f_2 + \cdots + f_k$. The frequency of x_1 is f_1 means observation x_1 is repeated f_1 times. The sum of all x_1 observations will be f_1x_1. Similarly, sum of all x_2 observations will be f_2x_2 and so on.

$$Mean(\bar{x}) = \frac{sum\ of\ all\ observation}{total\ number\ of\ observation}$$

$$= \frac{f_1x_1 + f_2x_2 + \cdots + f_kx_k}{f_1 + f_2 + \cdots + f_k} = \frac{\sum f_ix_i}{\sum f_i} = \frac{\sum f_ix_i}{n}$$

Short cut method:

As done earlier for the raw data, an assumed mean (average) A can be suitably chosen and the deviation of values $x_1, x_2, \ldots, x_k$ can be taken from A. Thus we will have the values of deviation obtained by $d_i = x_i - A$.

$$Mean(\bar{x}) = A + \frac{\sum f_id_i}{\sum f_i} = A + \frac{\sum f_id_i}{n}$$

Further, if all deviations have a common factor c, we can further simplify the calculation by dividing all the deviation by c, $u_i = \frac{x_i - A}{c}$

$$Mean(\bar{x}) = A + \left(\frac{\sum f_iu_i}{\sum f_i} \times c\right) = A + \left(\frac{\sum f_iu_i}{n} \times c\right)$$

Illustration 3: From the following data of the marks obtained by 60 students of a class, calculate the Mean marks of students.

Marks	20	30	40	50	60	70
No. of Students	8	12	20	10	6	4

Answer:

Let the mark be denoted by *x* and no. of students by *f*.

Marks (x)	**No. of Students (f)**	fx
20	8	160
30	12	360
40	20	800
50	10	500
60	6	360
70	4	280
-	**60**	**2460**

$$(\bar{x}) = \frac{\sum f_i x_i}{\sum f_i} = \frac{8(20) + 12(30) + 20(40) + 10(50) + 6(60) + 4(70)}{8 + 12 + 20 + 10 + 6 + 4}$$

$$= \frac{160 + 360 + 800 + 500 + 360 + 280}{60} = \frac{2460}{60} = 41$$

Hence Mean mark of students is 41 Marks.

Illustration 4: The time (in minutes) taken by a bus to travel between towns on different days is shown in the following table.

Time (Min.)	110	113	120	122	126
No. of Days	**7**	**17**	**11**	**10**	**5**

Find the mean travel time.

Answer:

Here values of the observation (time) is large. We will take assume mean A = 120.

Time (Minutes) x	No. of Days f	$d = x - 120$	df
110	7	-10	-70
113	17	-7	-119
120	11	0	0
122	10	2	20
126	5	6	30
-	**50**	-	**-139**

$$\mathrm{M}(\bar{x}) = A + \frac{\sum f_i d_i}{\sum f_i} = 120 + \frac{7(-10) + 17(-7) + 11(0) + 10(2) + 5(6)}{\mathbf{7 + 17 + 11 + 10 + 5}}$$

$$= 120 + \frac{(-70) + (-119) + 0 + 20 + 30}{\mathbf{50}} = 120 + \frac{-139}{\mathbf{50}} = 120 + (-2.78) = 117.22$$

Hence, mean travel time is 117.22 min.

Illustration 5: The price of an item changes from shop to shop. The following data are available. Find the mean price.

Price	206	212	218	220	224	230
No. of Shops	**1**	**3**	**14**	**9**	**8**	**5**

Answer:

As the observation values are large, we will calculate the mean using shortcut method. In which we select A=220, after calculating deviation we also found that the common factor of derived deviation is 2. Hence highest common factor of derived deviation is c=2.

Price x	No. of Shops f	$x-220$	$u=\frac{x-220}{2}$	uf
206	1	-14	-7	-7
212	3	-8	-4	-12
218	14	-2	-1	-14
220	9	0	0	0
224	8	4	2	16
230	5	10	5	25
-	**40**	**-**	**-**	**8**

$$\text{M }(\bar{x}) = A + (\frac{\sum f_i u_i}{\sum f_i} \times c) = 220 + (\frac{1(-7)+3(-4)+14(-1)+9(0)+8(2)+5(5)}{1+3+14+9+8+5} \times 2)$$

$$= 220 + (\frac{(-7)+(-12)+(-14)+0+16+25}{40} \times 2)$$

$$= 220 + (\frac{8}{40} \times 2) = 220 + (0.2 \times 2) = 220 + 0.4 = 220.4$$

Mean price of item is 220.4.

For Continuous frequency distribution:

When we transform the data into a continuous frequency distribution, each frequency shows the number of observations in that class. Thus we do not know each observation of that class. The mid-value of that class is taken as a representative for all the values in that class.

For example, let us consider a class 0-5 with frequency 7. Exact values of these 7 observations are not known. Hence the mid-value 2.5 is assumed for all 7 observations of this class which actually can be any number from 0 to 5.

Considering the mid-value of each class as x, the mean can be calculated by the same method as the one used for the discrete frequency distribution described earlier.

$$\text{Mid-Values } (x) = \frac{\textit{upper limit of the class}+\textit{lower limit of the class}}{2}$$

Illustration 6: The distribution of annual sales tax of different companies in a zone is given below. Find the mean sales tax of these companies.

Sales Tax (Thousands Rs.)	0-10	10-20	20-30	30-50	50-70
No. of Companies	**3**	**16**	**34**	**38**	**19**

Answer:

First we calculate mid-values of each class which is denoted as x. Later the mean can be calculated as the one used for the discrete frequency distribution described earlier.

Sales Tax (Thousands Rs.)	No. of Companies f	Mid-Values x	$x-25$	$u=\frac{x-25}{5}$	uf
0-10	3	5	-20	-4	-12
10-20	16	15	-10	-2	-32
20-30	34	25	0	0	0
30-50	38	40	15	3	114
50-70	19	60	35	7	133
-	**110**	**-**	**-**	**-**	**203**

In the above table we select assume mean A=25 and there is a common factor 5 in mid-values so c=5.

$$\text{M}\ (\bar{x}) = A + (\frac{\sum f_i u_i}{\sum f_i} \times c) = 25 + (\frac{3(-4) + 16(-2) + 34(0) + 38(3) + 19(7)}{3 + 16 + 34 + 38 + 19} \times 5)$$

$$= 25 + (\frac{(-12) + (-32) + (0) + 114 + 133}{110} \times 5)$$

$$= 25 + (\frac{203}{110} \times 5) = 25 + (1.85 \times 5) = 25 + 9.25 = 34.25$$

2.5 MEASURES CENTRAL TENDENCY: MEDIAN

We studied that mean is an appropriate average if we have data which are evenly distributed around the average and the data which do not have too large or too small values. It is said that mean does not become a good representative of data if these conditions are not satisfied. Another average is more suitable in such situation which is called as median. It is a positional average. In addition to median, quartiles, deciles, and percentiles are also other positional averages.

2.5.1 Meaning:

Median, quartiles, deciles, and percentiles are called positional averages because their values are found using the value of an observation at a specific position among the values of variable in the ordered data.

2.5.2 Definition of Median:

Median is defined as the value of middlemost observation when the data are arranged in either ascending or descending order. It is denoted by ***'M'***. Median value divides the data into two parts. In other words, 50% values of observations in the data are above the median and 50% observations have value less than the median.

2.5.3 Advantages and disadvantages of Median:

Advantages:

(1) It is easy to calculate and to understand.
(2) It can be found by inspection.
(3) It can be located by graph.
(4) It is only available average when the frequency distribution has open ended classes.
(5) It is less affected by too large or too small values.
(6) It can be calculated even if certain data are missing.

Disadvantages:

(1) It is rigidly defined.
(2) It is not based on all values.
(3) It is not suitable for further algebraic operations.
(4) It is less stable measure of central tendency as compared to mean.

2.5.4 Calculation of Median:

For ungrouped data:

As we have to find the value at the center, the observations have to be arranged in **ascending** or **descending** order.

For n arranged (in **ascending** or **descending** order) $x_1, x_2, \dots, x_n$ observation Median is found as follows:

$$Median\ (M) = \left(\frac{n+1}{2}\right)^{th} observation$$

For example, if we have 7 observation, the value of the $\left(\frac{7+1}{2}\right)^{th}$ i.e. the 4^{th} observation will be exactly the central value, which is called as Median.

Suppose the given data consist of 16 observations. Then as $\left(\frac{16+1}{2}\right)^{th}$ i.e. the 8.5, we say that the 8^{th} and 9^{th} observations are both in the center. In this case, median will be taken as the mean of these two (8^{th} and 9^{th} observation) central values.

Illustration 7: Find the Median for the following data showing runs scored by a batsman in his 20 innings.

32	28	47	63	71	9	60	10	96	14	31	148	53	67	29	10	62	40	80	54

Answer:

Let us arrange the data into ascending order.

9	10	10	14	28	29	31	32	40	47	53	54	60	62	63	67	71	80	96	148

Here, we have given runs of batsman of his 20 innings. So, n = 20.

Now,

$$Median\ (M) = \left(\frac{n+1}{2}\right)^{th} observation = \left(\frac{20+1}{2}\right)^{th} observation$$

$$= 10.5^{th}\ observation = \left(\frac{10^{th} obs.^{n} + 11^{th} obs.^{n}}{2}\right)$$

$$= \left(\frac{47+53}{2}\right) = \left(\frac{100}{2}\right) = 50\ Run$$

Hence, Median value of batsman runs is 50 run.

Illustration 8: There are 11 employees in an office. The monthly salaries (in Rs.) of the lowest paid 7 employees among them are 4500, 2100, 3400, 3600, 2500, 4200, 1500. What is the median monthly salary of all employees?

Some of the observations are missing here. We do not know the salaries of highest paid 4 employees. Suppose these values are *a*, *b*, *c*, *d* in their increasing order. These values are greater than the given observations as given salaries are of highest paid employees.

Now, we will arrange these data in ascending order.

1500, 2100, 2500, 3400, 3600, **4200**, 4500, a, b, c, d.

$$Median\ (M) = \left(\frac{n+1}{2}\right)^{th} observation = \left(\frac{11+1}{2}\right)^{th} observation = 6^{th} observation$$

$$Median\ (M) = 4200$$

Thus, the median salary of these employees is 4200 Rs.

For grouped data:

For discrete frequency distribution:

Suppose $x_1, x_2, \dots, x_k$ are the observation in the given data with frequencies $f_1, f_2, \dots, f_k$ respectively.

A frequency distribution generally shows the observations arranged in ascending order. We shall use cumulative frequencies to find the median for a frequency distribution where observations are arranged in an ascending order.

Here the median is found as follows:

$$Median\ (M) = value\ of\ \left(\frac{n+1}{2}\right)^{th}\ observation;$$

Where $n = total\ number\ of\ observation = \sum f_i$

Illustration 9: The following table shows the record of absent students of class during a month. Find the median of number of absent days per student.

Number of absent days of students	0	1	2	3	4	5
Number of students	**18**	**23**	**17**	**7**	**5**	**3**

Answer:

First we calculate the cumulative frequency.

Number of absent days of students	0	1	2	3	4	5	Total
Number of students (f)	**18**	**23**	**17**	**7**	**5**	**3**	**73**
Cumulative Frequency (cf)	**18**	**41**	**58**	**65**	**70**	**73**	**-**

Here, n = 73.

$$Median\ (M) = value\ of\ \left(\frac{n+1}{2}\right)^{th}\ observation$$

$$= value\ of\ \left(\frac{73+1}{2}\right)^{th}\ observation$$

$$= value\ of\ 37^{th}\ observation$$

It can be known from the cumulative frequencies that 19^{th} to 41^{st} observation have value 1. Hence, the 37^{th} observation has value 1.

$$\therefore Medain\ (M) = 1\ day$$

Illustration 10: The time required for typing a report by different typing is given in the following data. Find the median typing time using it.

Time for typing (min.)	10	11	12	13	14
Number of typists	**5**	**17**	**9**	**7**	**6**

Answer:

First we calculate the cumulative frequency.

Time for typing (min.)	10	11	12	13	14	Total
Number of typists	**5**	**17**	**9**	**7**	**6**	**44**
Cumulative Frequency ($\boldsymbol{cf}$)	**5**	**22**	**31**	**38**	**44**	**-**

Here, n = 44.

$$= value\ of\ (\frac{44+1}{2})^{th}\ observation$$

$$= value\ of\ 22.5^{th}\ observation$$

$$= \frac{value\ of\ 22^{nd} observation + value\ of\ 23^{rd} observation}{2}$$

It can be known from the cumulative frequencies that the 6^{th} to 22^{nd} observation have value 11 and the 23^{rd} to 31^{st} observation have value 12. Hence, the 22^{nd} and the 23^{rd} observation are 11 and 12 respectively.

$$Median\ (M) = \frac{11+12}{\mathbf{2}} = 11.5\ Min.$$

For continuous frequency distribution:

A continuous frequency gives the value of the variable in the form of class intervals and they are generally arranged in ascending order. In such cases, we will use the cumulative frequencies to find the median. These cumulative frequencies will show us the class containing median. For this, we take Median class = class containing the $(\frac{n}{2})^{th}$ observation.

$$\boldsymbol{Median\ (M) = L + (\frac{\frac{n}{2} - cf}{f} \times c)}$$

Where $n = total\ number\ of\ observation = \sum f_i$

$L = lower\ boundary\ point\ of\ the\ median\ class$

$cf\ = cumulative\ frequency\ of\ the\ class\ prior\ ro\ median\ class$

$f\ = frequency\ of\ the\ Median\ class$

$c = length\ of\ Median\ class$

Illustration 11: The monthly income (in thousand Rs.) of 130 persons living in a certain area is as follows:

Income (thousand Rs.)	Less than 4	4 – 8	8 – 12	12 – 20	20 – 28	28 – 36
No. of person	6	14	31	35	28	16

Find the median of income.

Answer:

Income (thousand Rs.)	Less than 4	4 – 8	8 – 12	12 – 20	20 – 28	28 – 36
No. of person (f)	6	14	31	35	28	16
Cumulative Frequency (cf)	**6**	**20**	**51**	**86**	**114**	**130**

Here, n = 130. So, (n/2) = 65. Hence 12 – 20 is median class because that class has cumulative frequency more than (n/2).

$$n = 130;\ L = 12;\ cf = 51;\ f = 35;\ c = 8$$

$$\boldsymbol{Median\ (M) = L + (\frac{\frac{n}{2} - cf}{f} \times c)}$$

$$\boldsymbol{Median\ (M) = 12 + (\frac{\frac{130}{2} - 51}{35} \times 8) = 12 + (\frac{65 - 51}{35} \times 8)}$$

$$\boldsymbol{= 12 + (\frac{14}{35} \times 8) = 12 + (0.4 \times 8) = 12 + 3.2 = 15.2}$$

Median income is 15.2 thousand Rs.

Illustration 12: The data about monthly expenditure on petrol for 75 families is given in the following table. Find the median expenditure on petrol for these families.

Expenditure (Rs.) on petrol	Upto 200	Upto 400	Upto 600	Upto 800	Upto 1000	Upto 1200
No. of families	2	8	17	32	57	75

Answer:

We are given the cumulative frequencies. We will find the frequency distribution.

Expenditure (Rs.) on petrol	**Upto 200**	**200 – 400**	**400 – 600**	**600 – 800**	**800 – 1000**	**1000 – 1200**
No. of families	2	6	9	15	25	18
Cumulative Frequency (cf)	2	8	17	32	57	75

Here, n = 75. So, (n/2) = 37.5. Hence 800 – 1000 is median class because that class has cumulative frequency more than (n/2).

$$n = 75;\ L = 800;\ cf = 32;\ f = 25;\ c = 200$$

$$\boldsymbol{Median\,(M) = L + (\frac{\frac{n}{2} - cf}{f} \times c)}$$

$$\boldsymbol{Median\,(M) = 800 + (\frac{\frac{75}{2} - 32}{25} \times 200) = 800 + (\frac{37.5 - 32}{25} \times 200)}$$

$$\boldsymbol{= 800 + (\frac{5.5}{25} \times 200) = 800 + (\frac{1100}{25}) = 800 + 44 = 844}$$

Thus, the median monthly expenditure on petrol of these families is 844 Rs.

Illustration 13: The level of air pollution (in ppm) in a city of different days is as follows. Find the median level of pollution.

Level of pollution (in ppm)	**250 & above**	**270 & above**	**290 & above**	**310 & above**	**320 & above**	**330 & above**	**340 & above**
No. of days	150	133	108	76	41	20	7

Answer:

A cumulative frequency distribution of ‘more than’ type is given here. We shall obtain the frequency distribution as well as the cumulative frequency distribution of ‘less than’ from it.

Level of pollution (in ppm)	**250 – 270`**	**270 – 290**	**290 – 310**	**310 – 320**	**320 – 330**	**330 – 340**	**340 & above**
No. of days	17	25	32	35	21	13	7
Cumulative Frequency (cf)	17	42	74	109	130	143	150

Here, n = 150. So, (n/2) = 75. Hence 310 – 320 is median class because that class has cumulative frequency more than (n/2).

$$n = 150;\ L = 310;\ cf = 74;\ f = 35;\ c = 10$$

$$\boldsymbol{Median\,(M) = L + (\frac{\frac{n}{2} - cf}{f} \times c)}$$

$$\boldsymbol{Median\ (M) = 310 + \left(\frac{\frac{150}{2} - 74}{35} \times 10\right)}$$

$$= \boldsymbol{310 + \left(\frac{75 - 74}{35} \times 10\right) = 310 + \left(\frac{1}{35} \times 10\right) = 310 + \frac{10}{35} = 310 + 0.29}$$

$$\boldsymbol{Median\ (M) = 310.29}$$

The, median for level of pollution 310.29 ppm.

2.5.5 Other positional averages:

We saw that the median divides the data in two equal parts. Sometimes we require values that divide data in more parts. We shall now study a few such positional averages.

Quartiles:

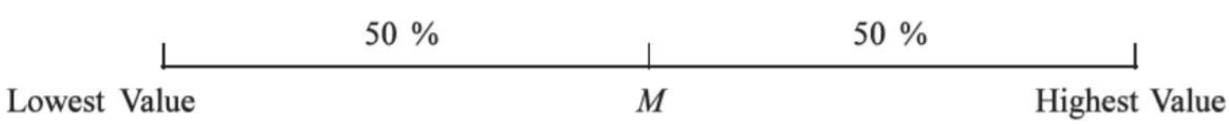
When the observations of the given data are arranged in ascending order, three values which divide the data in four equal parts are called quartiles. These three quartiles are denoted by Q_1, Q_2, and Q_3 respectively.

First 25% values of the data will be less than or equal to Q_1, next 25% values of the data will be between Q_1 and Q_2 and the further 25% values will be between Q_2 and Q_3. Hence we have 50% data values lying between Q_1 and Q_3, whereas 25% observations have value above Q_3.

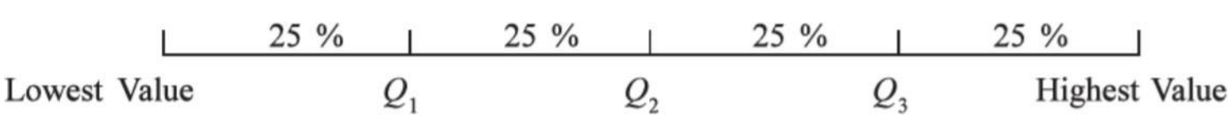
We can say that the j^{th} quartile Q_j, will divide the data such that 25j% observations will be below Q_j. (j = 1, 2, 3)

Thus, quartile Q_2, will have (25 × 2) % or 50% observations below Q_2. Hence Q_2 =Median=M.

For Median:



For Quartile:



Deciles:

Suppose the observations are arranged in the ascending order. Nine values which will divide the data in 10 equal parts are called as deciles which are denoted by D_1, D_2, …, D_{10}, respectively. 10%

observations will have value less than D_1, 20% observations will have value less than D_2, and so on. Thus, 10 j % observations will have value less than the j^{th} decile D_j (j = 1, 2, …, 9). We can see that $D_5 = M = Q_2$.

Percentiles:

Suppose the observations are arranged in the ascending order. The 99 values which divide the data in 100 equal parts are called percentiles which are denoted by $P_1, P_2, \ldots, P_{99}$ respectively. Here, 100 j % observations will have value less than the j^{th} percentile P_j (j = 1, 2, ,99). We can see that $D_1 = P_{10}$, $D_2 = P_{20}$, ..., $D_9 = P_{90}$. Similarly, $Q_1 = P_{25}$ and $Q_3 = P_{75}$. Moreover, $M = Q_2 = D_5 = P_{50}$.

Since median, quartiles, deciles, and percentiles are all positional averages, the calculation will have similar method. The following table shows the formula for finding the j^{th} quartile Q_j, the j^{th} decile D_j and the j^{th} percentile P_j.

Types of Data	Quartiles (j = 1, 2, 3)	
Raw Data and Discrete Frequency Distribution	$Q_j = value\ of\ the\ j\left(\frac{n+1}{4}\right)^{th}\ obs^n$	
Continuous Frequency Distribution	Class of $Q_j = j\left(\frac{n}{4}\right)^{th}$ obs^n. $Q_j = L + \left(\frac{j\left(\frac{n}{4}\right) - cf}{f} \times c\right)$	L=lower boundary point of the Quartile class f=frequency of the Quartile class c=length of Quartile class cf=cumulative frequency of the class prior to Q_j class
	Deciles (j = 1, 2, …, 9)	
Raw Data and Discrete Frequency Distribution	$D_j = value\ of\ the\ j\left(\frac{n+1}{10}\right)^{th}\ obs^n$	
Continuous Frequency Distribution	Class of $D_j = j\left(\frac{n}{10}\right)^{th}$ obs^n. $D_j = L + \left(\frac{j\left(\frac{n}{10}\right) - cf}{f} \times c\right)$	L=lower boundary point of the Decile class f=frequency of the Decile class c=length of Decile class cf=cumulative frequency of the class prior to D_j class
	Percentiles (j = 1, 2, …, 99)	
Raw Data and Discrete Frequency Distribution	$P_j = value\ of\ the\ j\left(\frac{n+1}{100}\right)^{th}\ obs^n$	
Continuous Frequency Distribution	Class of $P_j = j\left(\frac{n}{100}\right)^{th}$ obs^n. $P_j = L + \left(\frac{j\left(\frac{n}{100}\right) - cf}{f} \times c\right)$	L=lower boundary point of the Percentile class f=frequency of the Percentile class c=length of Percentile class cf=cumulative frequency of the class prior to P_j class

Illustration 14: Find Q_3, D_8 and P_{40} for the following data showing sales done by 20 sales men in last six months.

32	**28**	**47**	**63**	**71**	**9**	**60**	**10**	**14**	**96**
54	**31**	**148**	**53**	**67**	**29**	**10**	**62**	**40**	**80**

Answer:

Let's arrange the data into ascending order.

9	**10**	**10**	**14**	**28**	**29**	**31**	**32**	**40**	**47**
53	54	60	62	63	67	71	80	96	148

$$Q_3 = value\ of the\ 3\left(\frac{n+1}{4}\right)^{th} observation$$

$$= value\ of the\ 3\left(\frac{20+1}{4}\right)^{th} observation$$

$$= value\ of the\ 3(5.25)^{h} observation$$

$$= value\ of the\ 15.75^{th} observation$$

$$= 15^{th} observation + 0.75(16^{th} obs.^{n} - 15^{th} obs.)$$

$$= 63 + 0.75(67 - 63) = 63 + 0.75(4) = 63 + 3 = 66$$

So, 3rd quartile is 66.

$$D_8 = value\ of\ the\ 8\left(\frac{n+1}{10}\right)^{th} observation = value\ of the\ 8\left(\frac{20+1}{10}\right)^{th} observation$$

$$= value\ of the\ 8(2.1)^{h} observation = value\ of the\ 16.8^{th} observation$$

$$= value\ of the 16^{th} observation + 0.8\ (17^{th} obs.^{n} - 16^{th} obs.)$$

$$= 67 + 0.8\ (71 - 67) = 67 + 0.8(4) = 67 + 3.2 = 70.2$$

So, 8th Decile is 70.2.

$$P_{40} = value\ of\ the\ 40\left(\frac{n+1}{100}\right)^{th} observation = value\ of the\ 40\left(\frac{20+1}{100}\right)^{th} observation$$

$$= value\ of the\ 40(0.21)^{h} observation\ = 8.4^{th} observation$$

$$= value\ of the\ 8^{th} observation + 0.4(9^{th} obs.^{n} - 8^{th} obs.)$$

$$= 32 + 0.4(40 - 32) = 32 + 0.4(8) = 32 + 3.2 = 35.2$$

So, 40^{th} Percentile is 35.2.

Illustration 15: We have the following information from a survey of 100 customers of a bank on their number of visits to the bank during a month. Using the following information find 1^{st} quartile, 4^{th} decile and 89^{th} percentile.

Number of visit	0	1	2	3	4	5	6
Number of Customers	12	22	40	15	6	4	1

Answer:

First we calculate cumulative frequency.

Number of visit	0	1	2	3	4	5	6
Number of Customers	12	22	40	15	6	4	1
Cumulative Frequency (*cf*)	12	34	74	89	95	99	100

1^{st} quartile,

$$Q_1 = value\ of\ \left(\frac{n+1}{4}\right)^{th}\ observation$$

$$= value\ of\ \left(\frac{100+1}{4}\right)^{th}\ observation$$

$$= value\ of\ \left(\frac{101}{4}\right)^{th}\ observation$$

$$= value\ of\ 25.25^{th} observation$$

$$= 25^{th} observation\ + 0.25(26^{th} observation - 25^{th} observation)$$

According to cumulative frequency 25^{th} & 26^{th} observations are 1. So,

$$Q_1 = 1 + 0.25(1 - 1) = 1 + 0.25(0) = 1 + 0$$

$$Q_1 = 1$$

4^{th} Decile,

$$D_4 = \ value\ of 4\left(\frac{n+1}{10}\right)^{th}\ observation$$

$$= value\ of 4\left(\frac{100+1}{10}\right)^{th}\ observation$$

$$= value\ of\ 4\left(\frac{101}{10}\right)^{th}\ observation$$

$$= value\ of\ 4(10.1)^{h} observation$$

$$= value\ of\ 40.4^{th} observation$$

$$= 40^{th} observation + 0.4(41^{st} observation - 40^{th} observation)$$

According to cumulative frequency 40th & 41st observations are 2. So,

$$= 2 + 0.4(2 - 2) = 2 + 0.4(0) = 2 + 0$$

$$D_4 = 2$$

83rd Percentile,

$$P_{89} = value\ of\ 89\left(\frac{n+1}{100}\right)^{th}\ observation$$

$$= value\ of\ 89\left(\frac{100+1}{100}\right)^{th}\ observation$$

$$= value\ of\ 89\left(\frac{101}{100}\right)^{th}\ observation$$

$$= value\ of\ 89(1.01)^{h} observation$$

$$= value\ of\ 89.89^{th} observation$$

$$= 89^{th} observation + 0.89(90^{th} observation - 89^{th} observation)$$

According to cumulative frequency 89th observation is 3 & 90th observation is 4. So,

$$= 3 + 0.89(4 - 3) = 3 + 0.89(1) = 3 + 0.89$$

$$P_{89} = 3.89$$

Illustration 16: The following table shows data about the distance travelled (in km) by a salesman on different days. Find median Q_3, D_8 and P_{64} interpret it.

Distance travelled (km)	0 – 100	100 – 200	200 – 300	300 – 400	400 – 500	500 – 600
No. of Days	5	18	24	7	5	1

Answer:

Here we need to calculate cumulative frequency.

Distance travelled (km)	0 – 100	100 – 200	200 – 300	300 – 400	400 – 500	500 – 600
No. of Days	**5**	**18**	**24**	**7**	**5**	**1**
Cumulative Frequency (*cf*)	**5**	**23**	**47**	**54**	**59**	**60**

Here, n = 60. To calculate Q_3 i.e. third quartile we have to find $3\left(\frac{n}{4}\right) = 3\left(\frac{60}{4}\right) = 45$ to decide the Q_3 – Class. Hence, 200 – 300 is Q_3 – Class because that class has cumulative frequency more than $\frac{3n}{4}$.

$$L = 200, 3\left(\frac{n}{4}\right) = 45, cf = 23, f = 24, c = 100$$

$$Q_j = L + \left(\frac{\frac{jn}{4} - cf}{f} \times c\right)$$

$$Q_3 = L + \left(\frac{\frac{3n}{4} - cf}{f} \times c\right)$$

$$Q_3 = 200 + \left(\frac{3\left(\frac{60}{4}\right) - 23}{24} \times 100\right) = 200 + \left(\frac{45 - 23}{24} \times 100\right)$$

$$= 200 + \left(\frac{22}{24} \times 100\right) = 200 + (0.9167 \times 100) = 200 + 91.67 = 291.67$$

So, Q_3 = 291.67 that means 75% salesman travelled less than 291.67 km and remaining 25% salesman travelled more than 291.67 km.

Now to calculate 8th decile we find $8\left(\frac{n}{10}\right) = 8\left(\frac{60}{10}\right) = 48$. Here 300 – 400 class is D_8 class because that class has cumulative frequency more than $\frac{8n}{10}$.

$$L = 300, \frac{8n}{10} = 48, cf = 47, f = 7, c = 100$$

$$D_j = L + \left(\frac{j\left(\frac{n}{10}\right) - cf}{f} \times c\right)$$

$$D_8 = 300 + \left(\frac{8\left(\frac{60}{10}\right) - 47}{7} \times 100\right) = 300 + \left(\frac{48 - 47}{7} \times 100\right)$$

$$= 300 + \left(\frac{1}{7} \times 100\right) = 300 + \frac{100}{7} = 300 + 14.28 = 314.28$$

Here, $D_8 = 314.28$ that means 80% salesman travelled less than 314.28 km and remaining 20% salesman travelled more than 314.28 km.

To calculate 64[th] percentile we find $64\left(\frac{n}{100}\right) = 64\left(\frac{60}{100}\right) = 64(0.6) = 38.4$. Here 200 – 300 class is P_{64} class because that class has cumulative frequency more than $\frac{64n}{100}$.

$$L = 200, \frac{64n}{100} = 38.4, cf = 23, f = 24, c = 100$$

$$\boldsymbol{P_j = L + \left(\frac{j\left(\frac{n}{100}\right) - cf}{f} \times c\right)}$$

$$\boldsymbol{P_{64} = 200 + \left(\frac{64\left(\frac{60}{100}\right) - 23}{24} \times 100\right) = 200 + \left(\frac{38.4 - 23}{24} \times 100\right)}$$

$$\boldsymbol{= 200 + \left(\frac{15.4}{24} \times 100\right) = 200 + \frac{1540}{24} = 200 + 64.17 = 264.17}$$

Here, $P_{64} = 264.17$ that means 64% salesman travelled less than 264.17 km and remaining 36% salesman travelled more than 264.17 km.

2.6 MEASURE OF CENTRAL TENDENCY: MODE

We have earlier studied the mean and the median as the measures of central tendency. We shall now study ‘mode’ as another measure which is extensively used in business and commercial fields.

2.6.1 Meaning:

The value which gets repeated maximum number of times or the value occurring with maximum frequency in the given data is called as mode. It is denoted by Z or M_O.

It is very often used in business to give a representative value for a set of data. For example, see the following statements:

(1) On an average 3 languages are known to the students of this school.
(2) The average height of the men in our country is 1.7 m.
(3) The average daily production of our company is 50 items.
(4) The average daily overtime put in by the workers of a factory is 3 hours.

The value which is repeated most number of times is considered in the calculation of the average in these situations. As per the first statement, it is implied that most of the students know three languages. Thus we can say that mode is used as an average here.

2.6.2 Advantages and disadvantages of Mode:

Advantages:

(1) It is easy to understand and calculate.
(2) It can be found merely by inspection.
(3) It is not affected by too large or too small values.
(4) Its value can be found using graph.

Disadvantages:

(1) It is rigidly defined.
(2) There can be more than one mode for the given variable whereas sometimes the mode cannot be found.
(3) It is not based on all observation.
(4) It has less stability in sampling as compared to mean.
(5) It is not suitable for further mathematical calculations.

2.6.3 Calculation of mode:

For raw data and for discrete frequency distribution:

In these cases the mode can be found simply by inspection. We can find the mode as a value among the observations which is repeated maximum number of times or the one which has maximum frequency.

Illustration 17: The numbers of books purchased by each of the 15 persons from a book store are as follows.

1, 0, 2, 2, 3, 4, 2, 7, 2, 2, 5, 4, 2, 1, 2

Find the modal value for the number of books purchased.

Answer:

We can see that the value of 2 is repeated 7 times which is more than the number of repetitions of any value of the other observations. Hence mode Z or $M_O = 2$.

Thus, the mode of the number of books purchased is 2.

Illustration 18: TV sets assembled by a TV manufacturing company in a month are tested. The following table shows the numbers of defects per TV set. Find the mode for the number of defects.

Number of defects	0	1	2	3	4	5
Number of TV Sets	22	45	18	10	6	4

Answer:

An inspection of frequencies shows that the observation 1 has the maximum frequency 45.

Hence Z or M_O = 45 TV Sets

Thus the mode for the number of defects in TV sets is 1.

Illustration 19: The number of flights arriving at airport each hour during working hours of a day are recorded as follows.

3, 5, 4, 2, 7, 8

Find the mode for number of flights.

As all the values are appearing only once, we can't find the most common observation.

Hence the mode for the number of flights can’t be found from the given data using the definition.

For continuous frequency distribution:
When the data are converted into a frequency distribution with classes, the exact values of the observations are not available.

Similar to median, for mode also, the class containing mode is found first and the value of mode is found using it. The class having maximum frequency is called as modal class of the frequency distribution. The mode is further obtained using the following formula.

$$Mode\ (Z) = L + \left(\frac{f_m - f_1}{2f_m - f_1 - f_2} \times c\right)$$

Where L = lower boundary point of the modal class

f_m = frequency of the modal class

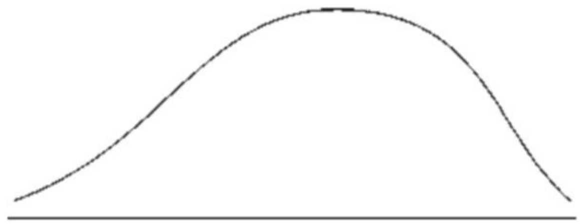
f_1 = frequency of the class prior to modal class

f_2 = frequency of the class succeeding to modal class

c = class length of the modal class

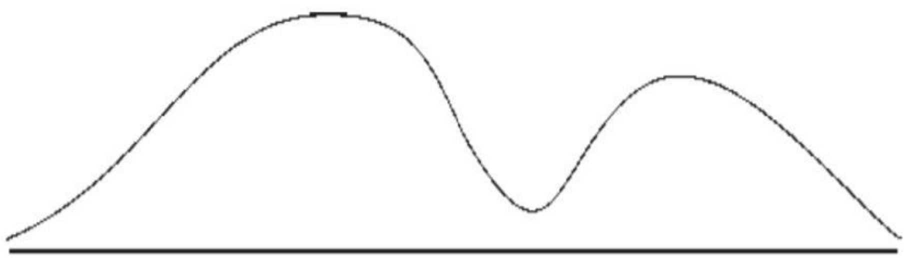
Note: The above formula can be used only if the distribution has classes of equal class length. Moreover, the formula can be used only in those cases where the maximum frequency is only for one class.

The frequency distribution in which the frequencies increase initially and then start decreasing after attaining the maximum frequency is called as a **regular frequency distribution**. Such distributions are also called as **unimodal distributions** as the distribution has only one mode. The frequency curve of such distributions is as follows:



Frequency curve of regular distribution

For binomial distribution, the frequencies increase and then decrease but then again increase and decrease. Such a frequency distribution is called as an **irregular frequency distribution** whose frequency curve is as follows.



Frequency curve of irregular distribution

Illustration 20: The following table shows data about the distance travelled (in km) by a salesman on different days. Find modal distance using the following data.

Distance travelled (km)	0 – 100	100 – 200	200 – 300	300 – 400	400 – 500	500 – 600
No. of Days	5	18	24	7	5	1

Answer:

Here, 200 – 300 class has maximum frequency. So, 200 – 300 class is defined as modal class.

$$L = 200, f_m = 24, f_1 = 18, f_2 = 7, c = 100$$

$$Mode\ (Z) = L + \left(\frac{f_m - f_1}{2f_m - f_1 - f_2} \times c\right)$$

$$= 200 + \left(\frac{24 - 18}{2(24) - 18 - 7} \times 100\right)$$

$$= 200 + \left(\frac{6}{48 - 25} \times 100\right)$$

$$= 200 + \left(\frac{6}{23} \times 100\right) = 200 + \left(\frac{600}{23}\right) = 200 + 26.09$$

$$Z = 226.09$$

2.6.4 Empirical formula for mode:

We have observed that mode is not well defined in many cases. The noted statistician Karl Pearson established a relation between mean, median and mode by studying their values for different data sets. He observed that for data that are not evenly distributed around average, difference between mean and mode is approximately 3 times the difference between mean and median i.e.,

$$(Mean - Mode) = 3\,(Mean - Median)$$

$$\bar{X} - Z = 3(\bar{X} - M)$$

$$\bar{X} - Z = 3\bar{X} - 3M$$

$$\bar{X} - 3\bar{X} + 3M = Z$$

$$Z = 3M - 2\bar{X}$$

This formula to find mode is called as an empirical formula because the value obtained from observation and not from the theory. The value of mode found using this formula can be negative if the frequency distribution is not evenly distributed around the average.

This formula for mode is used in the following situations:

- Each observation appears just once in raw data.
- More than one observation in a frequency distribution appears with highest frequency.
- The continuous distribution has classes of unequal length.
- The frequency distribution is a mixed distribution that is a part of it is discrete and the rest is continuous.
- The right or left end of the curve of the frequency distribution is too extended.

Illustration 21: The time between placing an order and its delivery was for a certain wholesaler as follows. Find the mode for this time.

Time (hours)	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50	50 – 55
No. of orders	2	5	7	5	6	7	3

Answer:

Here, maximum frequency is 7 for two different class 30 – 35 and 45 -50. So, we have to use empirical formula to find mode.

Time (hours)	No. of Orders (f_i)	Mid values (X_i)	$u_i = \frac{x_i - 37.5}{5}$	$f_i u_i$	Cumulative Frequency
20 – 25	2	22.5	-3	-6	2
25 – 30	5	27.5	-2	-10	7
30 – 35	7	32.5	-1	-7	14
35 – 40	5	37.5	0	0	19
40 – 45	6	42.5	1	6	25
45 – 50	7	47.5	2	14	32
50 – 55	3	52.5	3	9	35
	35			6	

Here, to calculate mean, A = 37.5 and c = 5.

$$\bar{x} = A + \left(\frac{\sum \boldsymbol{f_i u_i}}{\sum \boldsymbol{f_i}} \times c\right)$$

$$= 37.5 + \left(\frac{6}{35} \times 5\right) = 37.5 + \frac{30}{35} = 37.5 + 0.86$$

$$\bar{x} = 38.36$$

To, calculate median, (n/2) = (35/2) = 17.5. Hence class 35 – 40 is median class because that class has grater cumulative frequency compare to $\frac{n}{2}$.

$$L = 35, cf = 19, f = 5, c = 5, \frac{n}{2} = 17.5$$

$$M = L + \left(\frac{\frac{n}{2} - cf}{f} \times c\right) = 35 + \left(\frac{17.5 - 14}{5} \times 5\right) = 35 + (3.5)$$

$$M = 38.5$$

Now, using Empirical formula of Mode,

$$Z = 3M - 2\bar{x} = 3(38.5) - 2(38.36)$$

$$= 115.5 - 76.72$$

$$Z = 38.78$$

Thus, the mode for **time** between placing of an order and delivery is 38.78 hours.

Comparative study of mean, median, and mode

We have discussed advantages and disadvantages of various measures of central tendency. It is obvious from them that any particular average cannot be suitable for all types of practical problems. Each average has some specific applications and also has certain limitations.

Among all the averages, mean satisfies most of the requisites of a good average and hence it is used in most situations of data analysis. The most important property of mean is its compatibility for further algebraic computations. The advanced statistical methods applied for studying various characteristics of a population or for comparing two populations use mean as a representative value for the variable under consideration. These points make mean an optimum measure of central tendency.

However, mean cannot truly represent the entire set of data when the data are not evenly distributed around average. Many variables studied in agriculture, social sciences and in business activities are not found to be evenly distributed. Median is a better measure of central tendency in these situations. Median is used as an average in qualitative data like education, skill and consumer satisfaction.

Mode is particularly used in business and commerce fields. For qualitative data also, mode is found to be useful as an average. While deciding the dishes served at a restaurant and their flavor, the choice and taste of maximum customers is taken into consideration which is an example of Mode. Mode is extensively used for finding the average by readymade garment manufacturers and in foot wear industry.

Thus, the selection of average depends upon the following factors: (1) The nature of data (2) The nature of variable involved (3) The purpose of study (4) The type of classification used (5) The use of average for further statistical analysis

2.7 MEASURES OF DISPERSION

2.7.1 Meaning:

Dispersion is the extent to which values in a distribution differ from the average of the distribution. To quantify the extent of the variation, there are certain measures namely: (i) Range (ii) Interquartile range (iii) Absolute Mean Deviation (iv) Standard Deviation (v) Variance (vi) Coefficient of Variation.

2.7.2 Range:

Range (R) is the difference between the largest (L) and the smallest value (S) in a distribution. Thus, R = L – S Higher value of range implies higher dispersion and vice-versa.

2.7.3 Advantages and disadvantages of Range:

Advantages:

(1) The range is very clearly defined.
(2) Its computation is simple.
(3) Range is a useful measure especially when variability among the observations of the data is less.

Disadvantages:

(1) All observation of the data is not used in the computation of range.
(2) Range is very sensitive about sampling fluctuation.
(3) It is not a suitable measure for algebraic operations.
(4) It cannot be calculated for the frequency distribution having open-ended classes.

Illustration 22: We have given run of batsman of his last 5 matches. Find range of batsman’s run.

25, 46, 83, 39, 160

Answer: Here Range (R) = L – S = 160 – 25 =135. Range of batman’s run is 135 Run.

Illustration 23: We have the following information from a survey of 100 customers of a bank on their number of visits to the bank during a month. Find the range of Number of visit.

Number of visit	1	2	3	4	5	6
Number of Customers	22	20	15	16	14	13

Answer: Here in this illustration Maximum visit by customer are 6 visits and minimum visit by customer is 1 per month. So, Range (R) = L – S = 6 – 1 = 5 visit.

Illustration 24: The time between placing an order and its delivery was for a certain wholesaler as follows. Find the range for this time.

Time (hours)	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50	50 – 55
No. of orders	2	5	7	5	6	7	3

Answer: Here Largest Time is 55 hours and smallest time is 20 hours.

So, Range (R) = L – S = 55 – 20 = 35 Hours.

2.7.4 Interquartile Range:

The interquartile range describes the middle 50% of values when ordered from lowest to highest. Interquartile range is denoted as IQR. To find the interquartile range, first find the median (middle value) of the lower and upper half of the data. These values are 1st quartile Q_1 and 3rd quartile Q_3. The interquartile range is the difference between Q3 and Q1, i.e. IQR = $Q_3 - Q_1$.

Illustration 25: Find the interquartile range for the given data.

9, 5, 18, 12, 2, 15, 6, 19, 27, 1, 7

Answer:

To find quartile we arrange the data in to ascending order.

1, 2, 5, 6, 7, 9, 12, 15, 18, 19, 27

$$Q_1 = value\ of\ \left(\frac{n+1}{4}\right)^{th} observation = value\ of\ \left(\frac{11+1}{4}\right)^{th} observation$$

$$= value\ of\ 3^{rd}\ observation = 5$$

$$Q_3 = value\ of\ 3\left(\frac{n+1}{4}\right)^{th} observation = value\ of\ 3\left(\frac{11+1}{4}\right)^{th} observation$$

$$= value\ of\ 9^{th}\ observation = 18$$

Interquartile range,

$$IQR = Q_3 - Q_1 = 18 - 5 = 13$$

Illustration 26: Number of students solved MCQ out 20 is as follows. Find interquartile range.

No. of MCQ	2	6	10	14	18
No. of Students	90	150	100	200	80

Answer:

Let's calculate cumulative frequency.

No. of MCQ	2	6	10	14	18
No. of Students	90	150	100	200	80
Cumulative Frequency	90	240	340	540	620

1st Quartile, $(\frac{n+1}{4}) = (\frac{620+1}{4}) = 155.25$

$$Q_1 = value\ of\ 155.25^{th} observation$$

$$= value\ of\ 155^{th} observation + 0.25(156^{th} obs.^n - 155^{th} obs.)$$

Here, 155th and 156th observation is 6.

$$Q_1 = 6 + 0.25(6 - 6) = 6 + 0.25(0) = 6 + 0 = 6$$

3rd Quartile, $3\,(\frac{n+1}{4}) = 3\,(\frac{620+1}{4}) = 3\,(155.25) = 465.75$

$$Q_3 = value\ of\ 465.75^{th} observation$$

$$= value\ of\ 465^{th} observation + 0.25(466^{th} obs.^n - 465^{th} obs.)$$

Here, 465th and 466th observation is 14.

$$Q_3 = 14 + 0.75(14 - 14) = 14 + 0.75(0) = 14 + 0 = 14$$

Interquartile range,

$$IQR = Q_3 - Q_1 = 14 - 6 = 8$$

Illustration 27: For the following distribution of marks scored by a class of 40 students, Calculate Interquartile Range.

Marks	0 – 10	10 – 20	20 – 40	40 – 60	60 – 90
No. of Students	5	8	16	7	4

Answer:

Let's calculate cumulative frequency.

Marks	0 – 10	10 – 20	20 – 40	40 – 60	60 – 90
No. of Students	5	8	16	7	4
Cumulative Frequency	5	13	29	36	40

To calculate 1st Quartile, (n/4) = (40/4) = 10. So, 10 – 20 is Q_1 class because that class has grater cumulative frequency than (n/4).

$$Q_1 = L + (\frac{\frac{n}{4} - cf}{f} \times c) = 10 + (\frac{10 - 5}{8} \times 10) = 10 + (\frac{50}{8}) = 10 + 6.25 = 16.25$$

To calculate 3rd Quartile, (3n/4) = 3(40/4) = 30. So, 40 – 60 is Q_3 class because that class has grater cumulative frequency than (3n/4).

$$Q_3 = L + \left(\frac{\frac{3n}{4} - cf}{f} \times c\right) = 40 + \left(\frac{30 - 29}{7} \times 20\right) = 40 + \left(\frac{20}{7}\right) = 40 + 2.86 = 42.86$$

Interquartile Range, $(IQR) = Q_3 - Q_1 = 42.86 - 16.25 = 26.61$

2.7.5 Absolute Mean Deviation:

For raw data: Absolute mean deviation is an average of absolute deviation taken from mean.

For discrete & continuous frequency distribution: For frequency distribution absolute mean deviation is to be calculate by multiplying absolute deviation with its respective frequency of the class and divide its total by frequency.

2.7.6 Advantages and disadvantages of Absolute Mean Deviation:

Advantages:

(1) The mean deviation is a clearly defined measure of dispersion.
(2) It is superior measure to the range and the quartile deviation as all the observations are used in its computation.
(3) Its value is less affected by the extreme values (i.e. unduly the large and the small values) as compared to some other measures of dispersion.
(4) The absolute value of the difference between observation and the mean is used to measure the distance between an observation from the mean, which is an appropriate measure of distance.

Disadvantages:

(1) The computation of mean deviation is complicated as compared to the range and the quartile deviation.
(2) This measure is not suitable for algebraic operations.
(3) This measure is less used in advanced study of statistics as its definition is based on absolute value.
(4) It cannot be computed if the frequency distributions has open-ended classes.

Note: Mean deviation is frequently used to study the problems occurring in social sciences in general. It is also useful in Economics to determine economic inequality, in computing the distribution of personal wealth in the community or country, in forecasting weather and business cycles, etc.

2.7.7 Standard Deviation:

For raw data: Square root of average of square of deviation taken its mean is known as a standard deviation.

For discrete & continuous frequency distribution: Square root of average of square of deviation multiply with its respective frequency is known as standard deviation.

2.7.8 Advantages and disadvantages of Standard Deviation:

Advantages:

(1) Its definition is clear and precise.
(2) All the observations are used in its computation.
(3) Standard deviation is the most efficient measure of dispersion among all the measures of dispersion.
(4) Standard deviation is a suitable measure for algebraic calculations. For example, if the means and standard deviations of two data sets are given, the combined standard deviation of a new data set formed by combining the observations of two given data sets can be obtained. It is not possible to obtain a combined measure in case of other measures of dispersion by such an algebraic manipulation.
(5) Standard deviation is the most widely used measure of dispersion among all the measures of dispersion.

Disadvantages:

(1) The computation of standard deviation is more complicated as compared to computation of other measures of dispersion.
(2) The extreme observations get undue importance in the value this measure.
(3) It cannot be obtained if the frequency distributions have open-ended classes.

2.7.9 Variance:

For raw data, discrete & continuous frequency distribution: Square of standard deviation is known as Variance.

2.7.10 Formula to calculate Absolute Mean Deviation, Standard Deviation and Variance:

	Types of Data	
	For Raw Data	**For discrete & continuous frequency distribution**
Absolute Mean Deviation	$\boldsymbol{MD = \frac{\sum\lvert x - \bar{x}\rvert}{n}}$	$\boldsymbol{MD = \frac{\sum f\lvert x - \bar{x}\rvert}{\sum f}}$
Standard Deviation	$\boldsymbol{SD = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}}$	$\boldsymbol{SD = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f}}}$
Variance	$\boldsymbol{V = \frac{\sum(x - \bar{x})^2}{n}}$	$\boldsymbol{V = \frac{\sum f(x - \bar{x})^2}{\sum f}}$
	For non-integer Mean.	

Standard Deviation	$SD = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$	$SD = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$
Variance	$V = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$	$V = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$
	Short-cut Method to calculate Standard Deviation and Varinace. $u = x - A$ **or** $u = \frac{x-A}{c}$ **Where, A = Assumed Mean, c = common factor of** x	
Standard Deviation	$SD = \sqrt{\frac{\sum u^2}{n} - \left(\frac{\sum u}{n}\right)^2}$	$SD = \sqrt{\frac{\sum fu^2}{\sum f} - \left(\frac{\sum fu}{\sum f}\right)^2}$
Variance	$V = \frac{\sum u^2}{n} - \left(\frac{\sum u}{n}\right)^2$	$V = \frac{\sum fu^2}{\sum f} - \left(\frac{\sum fu}{\sum f}\right)^2$

2.7.11 Coefficient of Variation:

We have seen that the standard deviation is an absolute measure and it is expressed in terms of unit of the given observations of the data. Therefore, for the comparison of variability of two or more groups, their absolute measures cannot be used. For such a comparison, its relative measure, coefficient of standard deviation $\left(\frac{SD}{\bar{x}}\right)$ should be used. Often, the value of coefficient of standard deviation $\left(\frac{SD}{\bar{x}}\right)$ comes in fractional form, so Karl Pearson has suggested "Coefficient of Variation" as a relative measure which can be easily understood by common people. The coefficient of variation is obtained by multiplying coefficient of standard deviation by 100.

$$\text{Coefficient of Variation} = \frac{SD}{\bar{x}} \times 100$$

The coefficient of variation is measured in terms of percentage. i.e. coefficient of variation is percentage measure of standard deviation with respect to mean.

It is a very useful measure for comparing the dispersion of two or more data sets. A group of observations which has smaller value of coefficient of variation is said to be more stable and having less dispersion. Such a sequence is also said to be consistent from the point of view of variability. A sequence of observations which has larger value of coefficient of variation is said to be less stable and having more dispersion.

Illustration 28: The following data gives the monthly sales of 10 salesmen of particular brand water purifier. Find Mean Deviation, Standard Deviation, variance and coefficient of variation of sales using the data.

Sales (in Pc)	16	19	20	17	23	18	15	21	12	22

Answer:

											Total
Sales (in Pc)	16	19	20	17	23	18	15	21	12	22	**183**
$x - \bar{x}$	-2.3	0.7	1.7	-1.3	4.7	-0.3	-3.3	2.7	-6.3	3.7	**0**
$\lvert x - \bar{x} \rvert$	2.3	0.7	1.7	1.3	4.7	0.3	3.3	2.7	6.3	3.7	**27**
$(x - \bar{x})^2$	5.29	0.49	2.89	1.69	22.09	0.09	10.89	7.29	39.69	13.69	**104.1**

Mean,

$$\bar{x} = \frac{\sum x}{n} = \frac{183}{10} = 18.3$$

Absolute Mean Deviation,

$$MD = \frac{\sum \lvert x - \bar{x} \rvert}{n} = \frac{27}{10} = 2.7$$

Standard Deviation,

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = \sqrt{\frac{104.1}{10}} = \sqrt{10.41} = 3.23$$

Variance,

$$V = \frac{\sum (x - \bar{x})^2}{n} = \frac{104.1}{10} = 10.41$$

Coefficient of Variation,

$$CV = \frac{SD}{\bar{x}} \times 100 = \frac{3.23}{18.3} \times 100 = 17.65\%$$

Illustration 29: Find the absolute mean deviation, standard deviation, variance and coefficient of variation of the distribution of talk time (in minutes) per call:

Talk Time (minutes)	3	5	10	12	15
No. of Calls	4	7	6	2	1

Answer:

Talk Time (minutes) (x)	No. of calls (f)	fx	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$	$(x - \bar{x})^2$	$f\lvert x - \bar{x} \rvert$	$f(x - \bar{x})^2$
3	4	12	-4.3	4.3	18.49	51.6	221.88
5	7	35	-2.3	2.3	5.29	80.5	185.15
10	6	60	2.7	2.7	7.29	162	437.4
12	2	24	4.7	4.7	22.09	112.8	530.16
15	1	15	7.7	7.7	59.29	115.5	889.35
Total	20	146	-	-	-	522.4	2263.94

Mean

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{146}{20} = 7.3$$

Absolute Mean Deviation,

$$MD = \frac{\sum f|x - \bar{x}|}{\sum f} = \frac{522.4}{20} = 26.12$$

Standard Deviation,

$$SD = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f}} = \sqrt{\frac{2263.94}{20}} = \sqrt{113.197} = 10.64$$

Variance,

$$V = \frac{\sum f(x - \bar{x})^2}{\sum f} = \frac{2263.94}{20} = 113.197$$

Coefficient of Variation,

$$CV = \frac{SD}{\bar{x}} \times 100 = \frac{10.64}{7.3} \times 100 = 145.75\%$$

Illustration 30: Find the absolute mean deviation, standard deviation, Variance and coefficient of variation of the number of TV sets sold in last 16 months in a town.

No. of TV sets	10 – 30	30 – 50	50 – 70	70 – 90	90 – 110
No. of months	1	4	6	4	1

Answer:

No. of TV sets	No. of months (f)	(x)	fx	$x - \bar{x}$	$\|x - \bar{x}\|$	$(x - \bar{x})^2$	$f\|x - \bar{x}\|$	$f(x - \bar{x})^2$
10 – 30	1	20	20	-40	40	1600	40	1600
30 – 50	4	40	160	-20	20	400	80	1600
50 – 70	6	60	360	0	0	0	0	0
70 – 90	4	80	320	20	20	400	80	1600
90 – 110	1	100	100	40	40	1600	40	1600
Total	16	-	960		-	-	240	6400

Mean

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{960}{16} = 60$$

Absolute Mean Deviation,

$$MD = \frac{\sum f|x - \bar{x}|}{\sum f} = \frac{240}{60} = 4$$

Standard Deviation,

$$SD = \sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}} = \sqrt{\frac{6400}{60}} = \sqrt{106.67} = 10.33$$

Variance,

$$V = \frac{\sum f(x-\bar{x})^2}{\sum f} = \frac{6400}{60} = 106.67$$

Coefficient of Variation,

$$CV = \frac{SD}{\bar{x}} \times 100 = \frac{10.33}{60} \times 100 = 17.22\%$$

❖ CHECK YOUR PROGRESS

Answer the following Multiple Choice Questions

Que. 1 The observation which occurs most frequently in a sample is the
(A) median (B) mean deviation (C) standard deviation (D) mode

Que. 2 What is the median of the sample 5, 5, 11, 9, 10, 5, 8?
(A) 5 (B) 6 (C) 8 (D) 9

Que. 3 The following scores were obtained by eleven footballers in a goal-shoot competition: 5, 3, 6, 8, 7, 8, 3, 11, 6, 3, 6. The mean score was
(A) 3 (B) 6 (C) 8 (D) 11

Que. 4 The mean of ten numbers is 58. If one of the numbers is 40, what is the mean of the other nine?
(A) 18 (B) 60 (C) 162 (D) 540

Que. 5 The mean of 11 numbers is 7. One of the numbers, 13, is deleted. What is the mean of the remaining 10 numbers?
(A) 7.7 (B) 6.4 (C) 6.0 (D) 5.8

Que. 6 If the mean of 6 numbers is 41 than the sum of these numbers is
(A) 250 (B) 246 (C) 134 (D) 456

Que. 7 The difference of mode and mean is equal to
(A) 3(mean-median) (B) 2(mean-median) (C) 3(mean-mode) (D) 2(mean-mode)

Que. 8 If the mean is 11 and the median is 13 then the value of mode is
(A) 15 (B) 13 (C) 11 (D) 17

Que. 9 What is Geometric mean of observation 2, 8, 4, 16? (A) $\sqrt{1024}$ (B) $\sqrt{32}$ (C) $\sqrt{16}$ (D) $\sqrt{64}$

Que. 10 The distribution in which the values of median, mean and mode are not equal is considered as
(A) Experimental Distribution (B) Asymmetrical Distribution
(C) Symmetrical Distribution (D) Exploratory Distribution

Que. 11 If the value of three measures of central tendencies median, mean and mode then the distribution is considered as
(A) Negatively Skewed Modal (B) Triangular Modal (C) Unimoddel (D) Bimodel

Que. 12 In terms of dispersion difference, the measurement of dispersion for available data is classified as
(A) Average Measure (B) Distance Measure
(C) Average Deviation Measure (D) Availability Measures

Que. 13 The relative measures in measures of dispersion are also considered as
(A) Coefficient of Deviation (B) Coefficient of average
(C) Coefficient of variation (D) Coefficient of Uniformity

Que. 14 The categories of measures of dispersion are classified as
(A) Uniform measures (B) Relative Measures (C) Absolute Measures (D) Both B & C

Answer the following Question in one Sentence

Que. 1 Define Measures of Central Tendency.
Que. 2 Define Arithmetic Mean.
Que. 3 Write a formula of Arithmetic mean for Grouped Data (short cut method).
Que. 4 Define median.
Que. 5 Define mode.
Que. 6 Define relationship between Mean, Median and mode.
Que. 7 Write a formula of Geometric mean (for ungrouped data).
Que. 8 What is Positional Average?
Que. 9 What is Percentile?
Que. 10 Define Decile
Que. 11 What is Quartile?
Que. 12 What is Dispersion?
Que. 13 Give a definition of Range.
Que. 14 Define Standard Deviation.
Que. 15 What is Quartile Deviation?
Que. 16 Write formula of coefficient of Range.
Que. 17 What is coefficient of variation?

Answer the following Questions in detail

Que. 1 What is Average? State an objective of Averaging.
Que. 2 Write importance of Average.
Que. 3 What is mean? Give formula of mean for grouped and ungrouped data. Also write advantages and disadvantages of mean.
Que. 4 Explain median as Positional Average. Also state its advantages and disadvantages.
Que. 5 Describe Quartile, Deciles and Percentiles as Positional Average.
Que. 6 What is Mode? Write Empirical formula of Mode. State its advantages and disadvantages.
Que. 7 Describe comparative study of mean, median and mode.
Que. 8 What is measure of dispersion? State and describes different measures of dispersions.
Que. 9 What is Range? State its advantages and disadvantages.
Que. 10 What is Absolute mean deviation? State its advantages and disadvantages.
Que. 11 What is Standard Deviation? State its advantages and disadvantages.
Que. 12 The weekly growth (in cm.) in saplings grown in a nursery are:
1, 3.3, 1.9, 1.2, 2.4, 1.6, 1.3, 2.6, 1.6, 1.5.
Find the mean, median and mode of growth.

Que. 13 The following table gives the diameter (in mm.) of different screws selected from a large consignment. Find the mean, median and mode of diameter of screws.

Diameter of Screw (mm.)	30	35	40	45	50	55
No. of screws	4	10	15	8	5	3

Que. 14 The information of profits (in lakh Rs.) of 50 firms is given below. Find mean, median and mode of profit.

Profit (in lakh Rs.)	0-7	7-14	14-21	21-28	28-35
No. of firms	4	9	18	12	7

Que. 15 The distribution of demand of an item on different days is as follows. Find the mean, median and mode of demand.

Demand (units)	5-14	15-24	25-34	35-44	45-59	60-74	75-89
No. of days	4	17	12	18	16	13	10

Que. 16 Find Q_1, D_6, P_{40} and median for the following data showing runs scored by a batsman in his 20 innings.

32, 28, 47, 63, 71, 9, 60, 10, 96, 14, 31, 148, 53, 67, 29, 10, 62, 40, 80, 54

Que. 17 Find all qurtiles, D_3 and P_{62} for the data given below about marks scored by 15 students in class test.

8, 6, 7, 0, 2, 4, 6, 5, 5, 4, 8, 9, 3, 6, 7

Que. 18 The following tables shows production (in tons) of an item on different days. Find median, Q_3, D_8 and P_{83}.

Production (in tons)	0-100	100-200	200-300	300-400	400-500	500-600
No. of days	5	18	24	7	5	1

Que. 19 The following tables gives ages of 80 students selcted from a college. Find median age. Also find Q_1, D_4 and P_{32} for age of students.

Age (in years)	17	18	19	20	21	22	23
No. of students	22	28	44	30	16	12	8

Que. 20 The marks scored by a students in his last 10 tests of maths subject are 93, 81, 25, 48, 75, 37, 52, 72, 18 and 60. Find the range and the coefficient of range of his marks.

Que. 21 A bus company has 77 buses for travelling in the city. The information of number of passengers in bus on a particular day at a particular time is given below. Find the range and coefficient of range of number of passengers.

No. of passengers	2	7	10	18	25	30	37
No. of Buses	1	4	11	17	23	16	5

Que. 22 The frequency distribution of daily income (in thousands) of 80 shops of an area is as follows. Find the absolute and the relative measures of range of daily income from it.

Daily income (in thousand)	5-9	10-14	15-19	20-24	25-29	30-34
No. of shops	11	20	17	13	12	7

Que. 23 The measurements of weight (in kg) of 8 students of a class of a school are 48, 58, 60, 43, 75, 66, 51 and 81. Find the absolute mean deviation, coefficient of mean deviation, standard deviation and variance of weight of students.

Que. 24 The following distribution of closing prices (in Rs.) of shares of 100 different small scale industries on a certain day. Find the mean deviation, standard deviation and variance of the closing price of shares.

Price (in Rs.)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of industries	3	8	15	20	25	10	9	6	4

Que. 25 The information of the marks of 200 students of under graduate student studying in a college is given below. Find mean deviation and standard deviation of the mark of the students.

Marks	Less than 10	Less than 20	Less than 30	Less than 40	Less than 50	Less than 60	Less than 70
No. of students	5	17	47	92	142	179	200

UNIT-3 MATHEMATICAL MODEL

3.1 INTRODUCTION

Mathematical modeling is a process of obtaining a mathematical representation of some event or observations in the given subject / problem / system. It is a process that converts the observations in symbolic statements. During the process of building a mathematical model, the model will decide what factors are relevant to the given problem. Once a model has been developed it can be used to answer questions of the given problem. Generally the success of a model depends on how easily it can be used and how accurate are its predictions.

In context of management, a model can be defined as a set of mathematical or theoretical description of various variables of a system representing some aspects of a problem on some subject of interest. The model enables to conduct a number of experiment involving theoretical subjective manipulations to find some optimum solution to the problem on hand.

3.2 IMPORTANCE

Operational manager or management of an organization/industry needs to make decisions regarding man power, production, inventory, transportation, assignment of work, etc. depending on the data available. Through a model, one can explain the aspects of such problem/system. Mathematical model of a problem allow them to estimate the quantitative behavior of the problem. The obtained quantitative results can easily be compared with observational data to determine the strength and weakness of the model. Thus, mathematical model helps to management in determine the policies and actions.

3.3 TYPES OF MODELS

A manager/analyst has to study the problem, formulate the problem, identify the variables and formulate a model and select an appropriate technique to get optimal solution. Some of the basic models which are important for management of organizations are briefly described below:

1. Linear Programming Model

Basically, linear programming is a mathematical modeling technique in which an objective function is optimized (either maximized or minimized) subject to given various constraints. The objective function is to be maximized if it represents profit or sales and it is to be minimized if it represents loss or cost or time. Objective function is linear in nature. Constraints represent restrictions in resources and mostly they are represented by mathematical inequalities. This technique is useful to management for making quantitative decisions in their business planning.

2. Allocation Models (assignment problems)

Distribution of resources among the available alternatives in order to maximize return/sales or minimize cost/time is called allocation problem. In such problems the amount of each available resource is provided and the set of jobs to be done with their consumption of resources is also provided. The problem is to determine how much amount of each resource is to be allocated to each job. The transportation problem is an example of allocation model. An assignment problem is a special case of transportation problem as the number of resources and number of jobs is equal.

3. Inventory Models

Inventory model is a mathematical model that helps management to decide optimum level of inventories that should be preserved in a process of production. It also helps in making decisions about reordering level of materials in different situation of demand and supply.

4. Waiting Line Models (Queuing theory)

The waiting line model is suitable to use for service provider business where customer arrives and demands some service. Customer has to wait if the service provider/server is busy with other customer’s demand. Vehicle service provider, dental clinic, bank cashier are some example of waiting line model. This model help the management in balancing between the cost associated with waiting and the cost of providing service facilities.

5. Competitive Models (Game theory)

There are certain situations where two or more industrial or business units are involved in decision making under conflict situation. This means that decision-making is done to maximize the benefits and minimize the losses. The decision making much depends on the decision made or decision variables chosen by the opponent business organization. Such situations are known as competitive strategies. Competitive strategies are a type of business games. In business, competitive situations arise in advertising and marketing campaigns by competing business firms. When two or more opponents are involved under conditions of conflict or competition, it is known as game. The competitors in the game are called players. Not all but many of the competitive problems can be analyzed with the game theory.

6. Network Models (PERT/CPM techniques)

When the project is extremely large and complex such as construction of highway or residential building, it is important that managers must have a planning and scheduling system to produce the best outcomes and sometimes they also required to make necessary changes in the plan and schedule. Network technique is used to monitor and control such large projects and activities involved in the project. Two techniques PERT (Project Evaluation and Review Technique) and CPM (Critical Path Method) are most valuable techniques in managing the projects. PERT is used when exact time consumed by activities is not known, while CPM is used when the precise time of each activity involved in the project is known.

7. Simulation Models

Simulation is normally used to review the current, or predict the future, performance of a business process. The concept is designed to help practitioners and business owners discover new ways to improve their business processes through the use of mathematical, statistical and other analytical methods.

Simulation typically uses statistical and computer modeling to investigate the performance of a business process either for a new situation or to improve an existing set of processes. By modeling different process scenarios and outcomes, companies can minimize the traditional risks associated with change management initiatives without having to make changes in a 'live' business environment where performance could adversely be affected.

8. Dynamic Programming Models

Dynamic programming is a technique used to solve a multi-stage decision problems and for making a sequence of interrelated decisions. It is useful whenever the problem has large number of decision variables. Unlike linear programming, in this model there is no any standard mathematical formulation of the problem. This model provides general type of approach to problem solving.

9. Markov-Chain Models

The Markov-Chain process is a stochastic process. This model is applicable for decision making in a problem which is described in sequence of events and probability of each event depends only on the previous event, not on how we arrive at the current event. Traffic equilibrium problem is the best example where this model can be applied.

3.4 RESOURCE LIMITATION (CONSTRAINTS) AND OPTIMIZATION

An analyst has to formulate a model for the problem on hand. The steps involved in formulation of the model are listed below:

Step 1: Identify the decision variables.

Step 2: Identify the Resources and constraints.

Step 3: Identify the objective of the problem.

Step 4: Formulate the model by establishing relationship between variables and constraints.

Consider the following statement:

A company manufactures two products A and B, by doing the process on three different machines $M\text{–}I$, $M\text{–}II$ and $M\text{–}III$. Each unit of A requires 1 hour on machine $M\text{–}I$, 4 hours on machine $M\text{–}II$ and 6 hours on machine $M\text{–}III$. Similarly, product B requires 2 hour, 7 hours and 5 hours on Machine $M\text{–}I$, $M\text{–}II$ and $M\text{–}III$ respectively. In the coming month, 80 hours of machine $M\text{–}I$, 170 hours of machine $M\text{–}II$ and 220 hours of machine $M\text{–}III$ is available for production. Each unit of A brings a profit of Rs 4/- and B brings Rs. 6 per unit. How much units of product A and product B are to be manufactured in the next month by the company for maximizing the profit?

Step 1: Identifying Decision Variables

The Company is manufacturing two products A and B. We have to find out how much unit of product A and how much unit of product B are to be manufactured in the next coming month. Let us assume that the company manufactures x unit of product A and y unit of product B. Hence x and y are the two decision variables in our problem. Product A and product B are using or consuming available resources and hence they are known as competing candidates.

Step 2: Identifying Resources and Constraints

There are three machines $M\text{–}I$, $M\text{–}II$ and $M\text{–}III$ on which the products are manufactured.These are known as resources. The capacity of machines in terms of machine hours available are the available resources. The competing candidates have to use these available resources, which are limited in nature. Now in the above statement, machine $M\text{–}I$ has available 80 hours, machine $M\text{–}II$ has available a capacity of 170 hours and that of machine $M\text{–}III$ is 220 hours. The products have to use these machine hours in required proportion. That is one unit of product A consumes one hour of machine $M\text{–}I$, 4 hours of machine $M\text{–}II$ and 6 hours of machine $M\text{–}III$. Similarly, one unit of product B consumes 2 hour of machine $M\text{–}I$, 7 hours of machine $M\text{–}II$

and 5 hours of machine $M\text{– } III$. These machine hours given are the available resources and they are limited in nature and hence they are constraints given in the statement.

Step 3: Identifying Objective of the Problem

The objective of the problem is to find how much of product A and product B is to be manufactured so that the total profit is maximized? That is maximization of the profit or maximization of the returns is the objective of the problem. For this in the statement it is given that the profit contribution of product A is Rs 4/- per unit and that of product B is Rs. 6/- per unit.

Step 4: Formulating the model by establishing relationship between variables and constraints

We have assumed that company manufactures x units of product A and y units of product B. As one unit of x consumes one hour on machine $M\text{– } I$ and one unit of y consumes 2 hour on machine $M\text{– } II$, the total consumption by manufacturing x units of product A and y units of product B is, $1x + 2y$ and this should not exceed available capacity of 80 hours. Hence the mathematical relationship in the form of mathematical model is $1x + 2y \leq 80$. This is for resource machine $M\text{– } I$. Similarly for machine $M\text{– } II$ and $M\text{– } III$ we can formulate the mathematical relationship respectively as $4x + 7y \leq 170$ and $6x + 5y \leq 220$. Therefore, the mathematical model for these resources are:

$$1x + 2y \leq 80$$

$$4x + 7y \leq 170 \text{ and}$$

$$6x + 5y \leq 220.$$

Similarly for objective function, as the company manufacturing x units of product A and y units of product B and the profit contribution of product A and product B are Rs.4/- and Rs 6/- per unit of product A and product B respectively, the total profit earned by the company by manufacturing x and y units is $4x + 6y$. This we have to maximize. Therefore objective function is Maximize $4x + 6y$. At the same time, we have to remember one thing that the company can manufacture any number of units or it may not manufacture a particular product, for example say $x = 0$. But it cannot manufacture negative units of x and y. Hence one more constraint is to be introduced in the model i.e. a non–negativity constraint. Hence the mathematical representation of the contents of the statement is as given below:

Maximize $Z = 4x + 6y$ OBJECTIVE FUNCTION

Subject to a condition (written as s.t.)

$$1x + 2y \leq 80$$

$$4x + 7y \leq 170$$ STRUCTURAL CONSTRAINTS

$6x + 5y \leq 220$ and

Both x and y are ≥ 0 NON-NEGATIVITY CONSTRAINT.

The above illustrated mathematical model is an example of linear programming model. The general mathematical model of linear programming problem is described below:

Find the values of decision variables $x_1, x_2, \ldots \ldots, x_n$ so as to

Optimize (Maximize or Minimize) $Z = \sum_{j=1}^{n} c_j x_j$ (Objective function)

Subject to the linear constraints,

$$\sum_{j=1}^{n} a_{ij} x_j \ (\leq, =, \geq)_i \quad ; \quad i = 1, 2, \ldots \ldots, m$$

and $x_j \geq 0$; $j = 1, 2, \ldots \ldots, n$ (Constraints with non-negativity condition)

3.5 REAL WORLD APPLICATIONS

Following are some applications found in quantitative management where the system needs tobe modeled mathematically.

1. **Finance and Accounting**

   Dividend policies, investment and portfolio management, auditing, balance sheet and cash flow analysis, claim and complaint procedure, public accounting, break-even analysis, capital budgeting, cost allocation and control, financial planning, establishing costs for by-products and developing standard costs.

2. **Marketing**
   Selection of product-mix, marketing and export planning, advertising, media planning, selection and effective packing alternatives, sales effort allocation and assignment, launching a new product at the best possible time, predicting customer loyalty.

3. **Purchasing, Procurement and Exploration**

   Optimal buying and reordering with or without price quantity discount, transportation planning, replacement policies, bidding policies, vendor analysis.

4. **Production Management**

   Facilities planning: Location and size of warehouse or new plant, distribution centres and retail outlets, logistics, layout and engineering design, transportation, planning and scheduling.

Manufacturing: Aggregate production planning, assembly line blending, purchasing and inventory control, employment, training, layoffs and quality control, allocating R & D budgets most effectively.

Maintenance and project scheduling: Maintenance policies and preventive maintenance, maintenance crew size and scheduling, project scheduling and allocation of resources.

5. **Personnel Management**

Manpower planning, wage/salary administration, designing organization structures more effectively, negotiation in a bargaining situation, skills and wages balancing, scheduling of training programmes to maximize skill development and retention.

6. **Techniques and General Management**

Decision support systems and MIS; forecasting, making quality control more effective, project management and strategic planning.

7. **Government**

Economic planning, natural resources, social planning and energy, urban and housing problems, military, police, pollution control.

❖ CHECK YOUR PROGRESS

- **Answer the following multiple choice questions.**

Que. 1 In operations management, the___________are prepared for the situation.

(a) diagrammatic models (b) mathematical models

(c) physical models (d) none of these

Que. 2 Which of the following is not the phase of operational management methodology?

(a) formulating a problem (b) constructing a model

(c) controlling the environment (d) obtaining optimum solution

Que. 3 For management, models can help to determine

(a) policies (b) actions

(c) both (a) and (b) (d) none of these

Que. 4 Which technique is used in finding a solution for optimizing a given objective, such as profit maximization under certain constraints?

(a) waiting line (b) simulation

(c) game theory (d) linear programming

Que. 5 What is the objective function in linear programming problems?

(a) a linear function in an optimization problem

(b) a constraint for available resource

(c) a set of non-negativity conditions

(d) an objective for research and development of a company

Que. 6 Which of the following model has no any standard mathematical formulation technique?

(a) linear programming (b) dynamic programming
(c) allocation (d) competitive

Que. 7 Which of the following is a part of formulation of the model?

(a) identifying variables (b) identifying constraints

(c) identifying objective (d) all of these

- **Answer the following questions in brief.**

Que. 1 What is mathematical modeling?

Que. 2 Write steps involved in formulation of a model.

Que. 3 Give full form of PERT and CPM.

Que. 4 Define the term "players" in view of game theory. Que.

5 What do you mean by optimization?

Que. 6 How mathematical model is useful to management?

Que. 7 Identify decision variables in the following objective function:

Maximize $Z = 3x_1 + 2x_2 + x_3$.

- **Answer the following questions in detail.**

Que. 1 Explain mathematical modeling with its importance to the management of an organization.

Que. 2 Explain linear programming model, allocation model and inventory model.

Que. 3 Explain queuing model, dynamic programming model and Markov-Chain model.

Que. 4 Explain competitive model and simulation model.

Que. 5 Discuss on applications where the system needs to be modeled mathematically.

Que. 6 How to formulate a model? Explain using example.

Unit – 4 LINEAR PROGRAMMING

4.1 INTRODUCTION

4.2 IMPORTANCE

4.3 FORMULATION OF LINEAR PROGRAMMING

- 4.3.1 STRUCTURE OF LINEAR PROGRAMMING MODEL
- 4.3.2 ASSUMPTIONS OF LINEAR PROGRAMMING MODEL
- 4.3.3 GENERAL MATHEMATICAL MODEL OF LINEAR PROGRAMMING PROBLEMS
- 4.3.4 STANDARD FORM OF LINEAR PROGRAMMING PROBLEMS
- 4.3.5 STEPS FOR LINEAR PROGRAMMING MODEL FORMULATION
- 4.3.6 PROBLEMS ON FORMULATION OF LINEAR PROGRAMMING PROBLEMS

4.4 GRAPHICAL SOLUTION METHOD

- 4.4.1 DEFINITIONS RELATED TO SOLUTION OF LINEAR PROGRAMMINGPROBLEMS
- 4.4.2 STEPS FOR SOLUTION OF LINEAR PROGRAMMING PROBLEM USING GRAPHICAL METHOD
- 4.4.3 PROBLEMS ON SOLUTION OF LINEAR PROGRAMMING PROBLEM USING GRAPHICAL METHOD

4.5 SIMPLEX METHOD

- 4.5.1 STEPS FOR SOLUTION OF LINEAR PROGRAMMING PROBLEM BY SIMPLEX METHOD (MAXIMIZATION CASE)
- 4.5.2 PROBLEMS ON SOLUTION OF LINEAR PROGRAMMING PROBLEM USING SIMPLEX METHOD

4.1 INTRODUCTION

Linear programming (LP) is a mathematical modeling technique useful for economic allocation of limited resources, such as labour, material, machine, time, warehouse space, capital energy, etc., to several competing activities, such as products, services, jobs, new equipment, projects, etc., on the basis of a given criterion of optimality.

Linear programming modeling technique is applied for solving real life decision problems. LP models consist of certain common properties and assumptions. The word "linear" refers to linear relationship among variables in a model. Thus, a given change in one variable will always cause a resulting proportional change in another variable. e.g. doubling the investment on a certain project will exactly double the rate of return. The word "programming" refers to modeling and solving a problem mathematically that involves the economic allocation of limited resources by choosing a particular course of action or strategy among various alternative strategies to achieve the desired objective.

4.2 IMPORTANCE

Linear programming helps in making decisions with more objective approach. It is useful in getting the optimum use of productive resources. It improves the quality of decisions. It also indicates how a decision-maker can employ his productive factors effectively by selecting and distributing (allocating) these resources. Linear programming techniques provide possible and practical solutions.

Linear programming technique can be applied in many areas. Some of the important application areas of linear programming are listed below:

1. Agricultural applications: Farm economics, Farm management, Agricultural planning.
2. Military applications: Transportation problem that maximizing the total tonnage of bombs on targets, minimizing the cost with required level of protection in the attack on enemy.
3. Production management: Product mix, Production planning etc.
4. Financial management: Profit planning, Portfolio selection
5. Marketing management: Media selection, Traveling – Salesman problem, etc.
6. Personnel management: Staffing problem, Determination of equitable salaries.

4.3 FORMULATION OF LINEAR PROGRAMMING MODEL

4.3.1 Structure of Linear Programming Model

The general structure of LP model consists of three basis components: (i) Decision variables (ii) The objective function and (iii) The constraints.

(i) Decision variables (activities):

We need to evaluate various alternatives (courses of action) for arriving at the optimal value of objective function. For this we pursue certain activities (also called decision variables) usually denoted by x_1, x_2, … … , x_n. The value of certain variables may or may not be under the decision maker's control. If values are under the control of the decision maker, then they are said to be controllable otherwise uncontrollable. In an LP model all decision variables are continuous, controllable and non-negative.

(ii) The objective function:

The objective (goal) function of each LP problem is expressed in terms of decision variables to optimize the criterion of optimality such as profit, cost, revenue, distance, etc. The general form of objective function:

Optimize (Maximize or Minimize) $Z = c_1x_1 + c_2x_2 + \dots\dots + c_nx_n$

where $c_1, c_2, \dots\dots, c_n$ are constants.

The optimal value of the objective function is obtained by the graphical method or simplex method.

(iii) The constraints:

There are always certain limitations on the use of resources, e.g. labour, machine, raw material, space, money, etc. Such constraints (limitations) must be expressed as linear equalities or inequalities in terms of decision variables. The solution of an LP model must satisfy these constraints.

4.3.2 Assumptions of Linear Programming Model

The following are the major assumptions of an LP model:

1. Certainity:

   In LP model, it is assumed that all its parameters such as resources, profit (or cost) contribution per unit of decision variable and consumption of resources per unit of decision variable must be known and constant.

2. Additivity:

   The value of the objective function and the total amount of each resource used (or supplied) must be equal to the sum of the respective individual contribution (profit or cost) of the decision variables. For example, the total profit earned from the sale of two products A and B must be equal to the sum of the profits earned separately from product A and product B.

3. Linearity (or proportionality):

   The amount of each resource used (or supplied) and its contribution to the profit (or cost) in objective function must be proportional to the value of each decision variable. For example, if production of one unit of a product uses 5 hours of a particular resource, then making 3 units of that product uses 15 hours of that resource.

4. Divisibility (or continuity):

   The solution values of decision variables are allowed to assume continuous values. For example, it is possible to collect 6.254 thousand liters of milk by a milk dairy and such variables are divisible. But it is not desirable to produce 2.5 machines and such variables are not divisible and therefore must be assigned integer values. Hence, if any of the variable can assume only integer values, LP model is no longer applicable.

4.3.3 General Mathematical Model of Linear Programming Problems

Find the values of decision variables $x_1, x_2, \dots \dots, x_n$ so as to

Optimize (Max. or Min.) $Z = c_1x_1 + c_2x_2 + \dots \dots + c_nx_n$ (Objective function)

Subject to the linear constraints,

$$
\begin{array}{l}
a_{11}x_1 + a_{12}x_2 + \dots \dots + a_{1n}x_n \ (\leq, =, \geq) \ b_1 \\
a_{21}x_1 + a_{22}x_2 + \dots \dots + a_{2n}x_n \ (\leq, =, \geq) \ b_2 \\
\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \dots \dots + a_{mn}x_n \ (\leq, =, \geq) \ b_m
\end{array}
$$

and $x_1, x_2, \dots \dots, x_n \geq 0$

4.3.4 Standard form of Linear Programming Problems

Standard form of Maximization Linear Programming Problem:

(i) Objective function Z is to be maximized
(ii) All decision variables $x_1, x_2, \dots \dots, x_n$ are non – negative (≥ 0)
(iii) All constraints are with inequality "less than or equal to" $(\leq)$.

Standard form of Minimization Linear Programming Problem:

(i) Objective function Z is to be minimized
(ii) All decision variables $x_1, x_2, \dots \dots, x_n$ are non – negative (≥ 0)
(iii) All constraints are with inequality "greater than or equal to" $(\geq)$.

4.3.5 Steps for Linear Programming Model Formulation:

Step 1: Identify the decision variables

Step 2: Identify the problem data

Step 3: Formulate the constraints

Step 4: Formulate the objective function.

4.3.6 Problems on Formulation of Linear Programming Problems

Illustration 4.1:

A vendor assembles and sells two items TV and computer. Both require two types of technician, hardware technician and software technician for completing process. One TV requires 4 hours of hardware technician's time and 2.5 hour of software technician's time. One computer requires 3 hour of hardware technician's time and 5 hours of software technician's time. The hardware technician has 250 hours and software technician has 200 hours of monthly available time. Profit on one TV is Rs. 800 and on one computer is Rs. 1000. Assuming that all the TVs and computers assembled are sold, how many TVs and computers should be assembling monthly to obtain the maximum profit?

Solution:

Decision variables:

Suppose number of monthly assembling of TVs = x

and number of monthly assembling of computers = y

The information provided in the problem can be summarized as below:

Item	No. of Units	Hardware Technician's Time (Hour)	Software Technician's Time (Hour)	Profit/Unit (Rs.)
TV	x	4	2.5	800
Computer	y	3	5	1000
Monthly Available Time		250	200	

LP model formulation

Objective function:

Maximize (total profit) $Z = 800x + 1000y$

Subject to the constraints:

$4x + 3y \leq 250$ Hardware technician time availability

$2.5x + 5y \leq 200$ Software technician time availability

$x, y \geq 0$ Non-negativity constraint

Illustration 4.2:

A furniture maker manufactures tables, chairs and sofa-sets. Each product requires two types of row material wood and other material. One table requires 3 units of wood and 4 units of other material. One chair requires 2 units of wood and 3 units of other material and the requirement for a sofa-set is 5 and 7 units of wood and other material respectively. Total available units of wood and other material is 4000 and 6000 respectively. A market survey indicates that the minimum demand of tables, chairs and sofa-sets is 400, 600 and 250 units respectively. Profit per table, chair and sofa-set is Rs. 200, 125 and 450 respectively. Formulate this problem as an LP model to determine the number of units of each product which will maximize profit.

Solution:

The data provided in the problem can be tabulated as below:

		Row material			
Item	No. of Units	Wood	Other material	Demand	Profit/Unit (Rs.)
Table	x	3	4	400	200
Chair	y	2	3	600	125
Sofa-set	z	5	7	250	450
Available material		4000	6000		

Decision variables:

Suppose x_1, x_2 and x_3 are number of tables, chairs and sofa-sets respectively to be manufactured.

LP model formulation

Objective function:

Maximize (total profit) $Z = 200x_1 + 125x_2 + 450x_3$

Subject to the constraints:

Row material requirement $3x_1 + 2x_2 + 5x_3 \leq 4000$

$4x_1 + 3x_2 + 7x_3 \leq 6000$

Market demand $x_1 \geq 400$

$x_2 \geq 600$

$x_3 \geq 250$

Note: Here non-negative constraints are not required as market demand is non-negative.

Illustration 4.3:

A company wants to advertise its product on TV and social media to reach different numbers of customers within three age-groups: over 60, between 40 and 60, and under 40 year old. One minute of TV commercial time costs Rs. 5000 and will reach an average of 13000 viewers in the over 60 group, 12000 customers in the 40 to 60 group, and 7000 in the under 40 group. One minute of social media time costs Rs. 2000 and will reach 2000 users in the over 60 age group, 10000 in the 40 to 60 age group, and 15000 in the under 40 group. The company wants to have a total exposure of 50000 in the over 60 group, 70000 in the 40-60 age-group, and 80000 in the under-40 group. Formulate an LP model to determine the amount of different commercial minutes to use at the minimum cost.

Solution:

Decision Variables:

Let number of minutes of TV commercials = x

and number of minutes of Social media commercials = y

LP model formulation

Objective Function:

Minimize (total cost) $\mathrm{Z} = 5000x + 2000y$

Subject to the constraints:

$13000x + 2000y \geq 50000$ viewers of >60 age group

$12000x + 10000y \geq 70000$ viewers of 40-60 age group

$7000x + 15000y \geq 80000$ viewers of <40 age group

$x, y \geq 0$ Non-negativity constraint

Illustration 4.4:

Following chart shows the different nutrient content per unit of four different food items. It also shows the cost per unit of each food item.

	Food Item			
Nutrient	A	B	C	D
Calorie	500	250	200	400
Carbohydrate (grams)	3	3	4	4
Protein (grams)	0	2	3	2
Cost (per unit) (Rs.)	35	15	20	50

Formulate a linear programming problem to minimize the diet cost in such a way that it contains at least 600 calories, 12 grams carbohydrate and 5 grams protein.

Solution:

Decision Variables:

Let the diet contain x_1 unit of food item A, x_2 unit of food item B, x_3 unit of food item C and x_4 unit of food item D.

LP model formulation

Objective Function:

Minimize (total cost) $\mathrm{Z} = 35x_1 + \ 15x_2 + 20x_3 + 50x_4$

Subject to the constraints:

$$500x_1 + 250x_2 + 200x_3 + 400x_4 \geq 600 \qquad \text{Calorie}$$

$$3x_1 + 3x_2 + 4x_3 + 4x_4 \geq 12 \qquad \text{Carbohydrate}$$

$$2x_2 + 3x_3 + 2x_4 \geq 5 \qquad \text{Protein}$$

$$x_1, x_2, x_3, x_4 \geq 0 \qquad \text{Non-negativity constraint}$$

4.4 GRAPHICAL SOLUTION METHOD

After obtaining the mathematical formulation of given linear programming problem (LPP), the appropriate method is applied for the optimum solution of the problem. The graphical solution method is suitable for the LP problem with two decision variables. This method cannot be applied to the LP problem with more than two decision variables.

4.4.1 Definitions Related to Solution of Linear Programming Problem

Solution: The set of values of decision variables x_j $(\mathrm{j} = 1, 2, \ldots \ldots, n)$ which satisfy the constraints of an LP problem is said to constitute solution to that LP problem.

Feasible solution: The set of values of decision variables x_j $(\mathrm{j} = 1, 2, \ldots \ldots, n)$ which satisfy all the constraints and non-negativity conditions of an LP problem simultaneously is said to constitute the feasible solution to that LP problem.

Infeasible solution: The set of values of decision variables x_j $(\mathrm{j} = 1, 2, \ldots \ldots, n)$ which do not satisfy all the constraints and non-negativity conditions of an LP problem simultaneously is said to constitute the infeasible solution to that LP problem.

Basic solution: For a set of m simultaneous equations in n variables $(n > m)$, a solution obtained by setting $(n - m)$ variables equal to zero and solving for remaining m equations in m

variables is called a basic solution.

The $(n - m)$ variables whose value did not appear in this solution are called non-basic variables and the remaining m variables are called basic variables.

Basic feasible solution: A feasible solution to an LP problem which is also the basic solution is called the basic feasible solution. i.e. all basic variables assume non-negative values. Basic feasible solutions are of two types:

Degenerate: A basic feasible solution is called degenerate if value of at least one basic variable is zero.

Non-degenerate: A basic feasible solution is called non-degenerate if values of all m basic variables are non-zero and positive.

Optimum basic feasible solution: A basic feasible solution which optimizes (maximizes or minimizes) the objective function value of the given LP problem is called an optimum basic feasible solution.

Unbounded solution: A solution which can increase or decrease the value of objective function of the LP problem indefinitely is called an unbounded solution.

4.4.2 Steps for Solution of Linear Programming Problem using Graphical Method

Step 1: Develop LP model.

State the given problem in the mathematical LP model.

Step 2: Plot constraints on graph paper and decide the feasible region.

Replace the inequality sign in each constraint by an equality sign. Draw these straight lines on graph paper and decide each time the area of feasible solutions according to the inequality sign of the constraint. Shade the common portion of the graph that satisfies all the constraints simultaneously. The final shaded area is called the feasible region of the given LP problem.

Step 3: Examine extreme points (corners) of the feasible region.

(a) Determine the coordinates of each corner point of the area of feasible region where all constraints overlap.
(b) Evaluate the value of the objective function at each extreme point.
(c) Determine the extreme point of the feasible region that has optimum objective function value.

4.4.3 Problems on Solution of Linear Programming Problem using Graphical Method

Illustration 4.5:

Solve the following LP problem using graphical method:

Maximize $Z = 4x + 3y$

Subject to $2x + y \leq 30$

$x + 2y \leq 40$

$x + y \leq 25$

$x, y \geq 0$

Solution:

Let us first determine the feasible region (solution space).

Since the two decision variables x and y are non-negative, consider only the first quadrant of XY-plane as shown in Fig. 1.

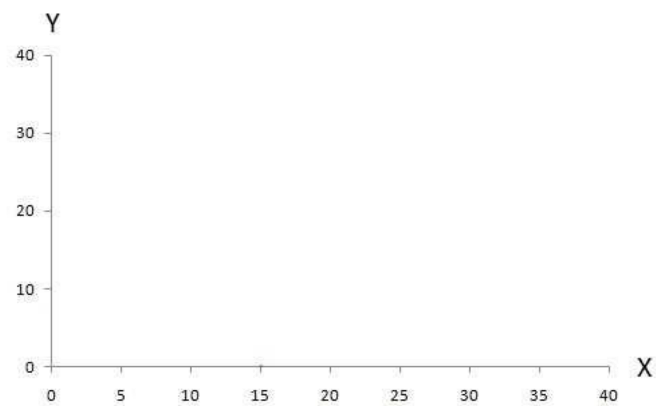


Figure 1

Consider first constraint i.e. $2x + y \leq 30$.

Treat this inequality as an equation $2x + y = 30$.

Put $y = 0$ in this equation, we get $x = 15$.

So we get coordinates of first point as (15, 0).

Similarly by putting $x = 0$, we get $y = 30$ and the coordinate (0, 30).

Plot these two points (15, 0) and (0, 30) on the graph and join them to get a straight line representing the equation $2x + y = 30$ as shown in Fig. 2. This line divides the first quadrant into two regions R_1 and R_2. A point (0, 0) falls in the region R_1 and the point (0, 0) satisfies the inequality $2x + y \leq 30$. Thus, the region R_1 including the points on the line is a region where the first constraint is satisfied.

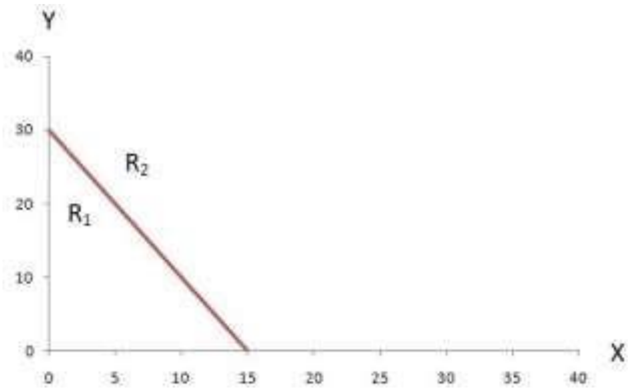


Figure 2

Now consider the second constraint $x + 2y \leq 40$.

Treat this inequality as an equation $x + 2y = 40$.

Putting $y = 0$, we get $x = 40$ and the coordinate (40, 0) and

by putting $x = 0$, we get $y = 20$ and the coordinate (0, 20).

So, the line representing this equation is a line obtained by joining two coordinates (40, 0) and (0, 20) as shown in Fig. 3. The inequality $x + 2y \leq 40$ is satisfied in the region R'_1.

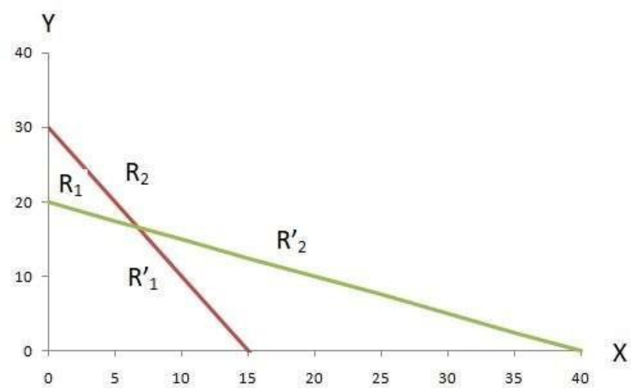


Figure 3

Similarly, draw a line which represents the equation $x + y = 25$ by joining two coordinates (25, 0) and (0, 25) and determine the region where the inequality $x + y \leq 25$ is satisfied.

The intersection of all regions, where the given constraints are satisfied, is the feasible region (The common region where all constraints are satisfied together). The feasible region is shown in Fig. 4 by the shaded area OABC.

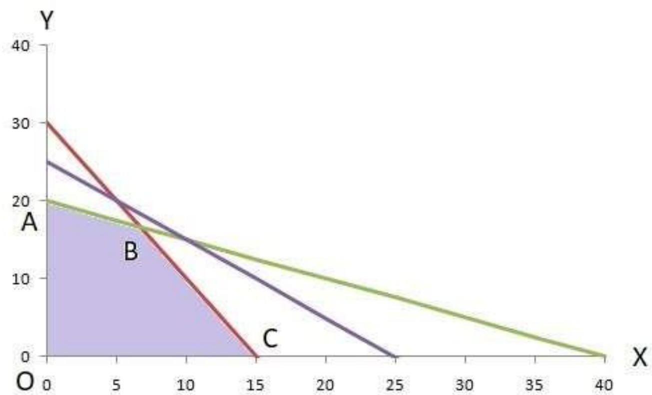


Figure 4

Since the optimal value of the objective function, if it exists, occurs at the vertices (corner points) of the feasible region, it is necessary to determine their coordinates. These coordinates can be obtained from the graph or by solving the equation of the lines.

By examining the graph in Fig. 4, it is clear that the coordinate of vertices O, A and C of the feasible region are (0, 0), (0, 20) and (15, 0) respectively. The coordinates of the vertex B, which is an intersection of two lines representing the equations $2x + y = 30$ and $x + 2y = 40$, can be obtained by solving these equations.

$$2x + y = 30 \ldots \ldots \ldots (1)$$

$$x + 2y = 40 \ldots \ldots \ldots (2)$$

Multiplying Eq. (2) by 2 and then subtracting Eq. (1), we get

$$3y = 50 \quad \text{or} \quad y = \frac{50}{3}$$

Substituting this value of y in Eq. (1), we get

$$x = \frac{20}{3}$$

Thus, the coordinate of B is $\frac{20}{3}\ \frac{50}{3}$

Computation of objective function value at each of this vertex is shown in the Table 4.1.

Table 4.1: Feasible solutions

Vertex	Coordinates	Value of objective function $Z = 4x + 3y$
O	(0, 0)	4(0) + 3(0) = 0
A	(0, 20)	4(0) + 3(20) = 60
B	$\left(\frac{20}{3}, \frac{50}{3}\right)$	$4(\frac{20}{3} + 3\left(\frac{50}{3}\right) = \frac{230}{3}$
C	(15, 0)	4(15) + 3(0) = 60

Since objective function Z is to be maximized, from Table 4.1 we can see that maximum value of Z is $\frac{230}{3}$ at $x = \frac{20}{3}$ and $y = \frac{50}{3}$. Hence the optimum solution to the given LP problem is: $x = \frac{20}{3}, y = \frac{50}{3}$ and Maximum $Z = \frac{230}{3}$.

Illustration 4.6:

Use graphical method to solve the following LP problem:

Minimize $\quad Z = 5x + 8y$

Subject to $\quad 25x + 20y \leq 300$

$$2x + y \geq 10$$

$$6x + 10y \leq 120$$

$$y \geq 5$$

$$x, y \geq 0$$

Solution:

In order to determine feasible region, plot each constraint on graph paper by first considering it as a linear equation. Then use inequality of each constraint and find the common area where all constraints are satisfied. This area is a feasible region which is shown in Fig. 5 as shaded area ABCD.

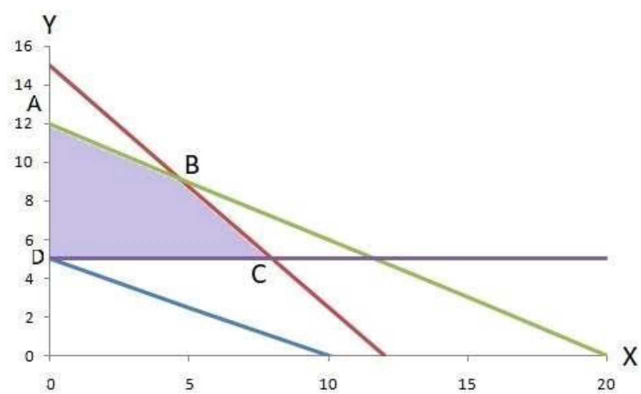


Figure 5

It is clear from the Fig. 5 that the coordinates of vertices A and D are (0, 12) and (0, 5) respectively. The vertex B is an intersection of two lines

$$25x + 20y = 300 \dots \dots \dots (1)$$

and $6x + 10y = 120 \dots \dots \dots (2)$

Multiplying Eq. (2) by 2 and then subtracting it from Eq. (1), we get

$$13x = 60 \; or \; x\frac{60}{13}$$

Substituting this value of x in Eq. (2), we get

$$y = \frac{120 - \frac{360}{13}}{10} = \frac{120}{13}$$

Thus, the coordinate of B is $(\frac{60}{13}, \frac{120}{13})$

Similarly, the coordinate of the vertex C, which is the intersection of two equations $y = 5$ and $25x + 20y \leq 300$, be obtained as (8, 5).

Computation of objective function value at each of this vertex is shown in the Table 4.2.

Table 4.2: Feasible solutions

Vertex	**Coordinates**	**Value of objective function** $Z = 5\text{x} + 8y$
A	(0, 12)	5(0) + 8(12) = 96
B	$(\frac{60}{13}, \frac{120}{13})$	$5(\frac{60}{13} + 8\left(\frac{120}{13}\right) = 96.92$
C	(8, 5)	5(8) + 8(5) = 80
D	(0, 5)	5(0) + 8(5) = 40

Since objective function Z is to be minimized, from Table 4.2 we can see that minimum value of Z is 40 at $x = 0$ and $y = 5$. Hence the optimum solution to the given LP problem is: $x = 0, y = 5$ and Min $Z = 40$.

Illustration 4.7:

Use graphical method to solve the following LP problem:

Minimize $Z = 20x + 40y$

Subject to $6x + y \geq 18$

$x + 4y \geq 12$

$2x + y \geq 10$

$x, y \geq 0$

Solution:

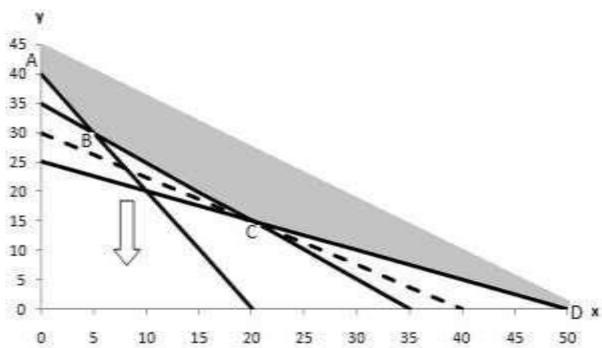


Figure 6

Plot each constraint on graph paper by first treating it as a linear equation. Then use inequality of each constraint and find the common area where all constraints are satisfied. This area is a feasible region which is shown in Fig. 6 as shaded area. The coordinates of vertices of the feasible region are: A=(0, 40), B=(5, 30), C=(20, 15), D=(50, 0). Computation of objective function value at each of this vertex is shown in Table 4.3.

Table 4.3: Feasible solutions

Vertex	**Coordinates**	**Value of objective function** $Z = 20\text{x} + 40y$
A	(0, 40)	20(0) + 40(40) = 160
B	(5, 30)	20(5) + 40(30) = 135
C	(20, 15)	20(20) + 40(15) = 120
D	(50, 0)	20(50) + 40(0) = 150

Since objective function Z is to be minimized, from Table 4.3 we can see that minimum value of Z is 120 at $x = 20$ and $y = 15$.

From Fig. 6 one can see that the feasible region is unbounded. So it may or may not be the optimal solution of the given problem. To determine, let us plot a straight line representing the objective function $3x + 4y = 120$. This line is shown as dashed line in Fig.6. The area where the inequality $3x + 4y < 120$ satisfied is indicated by arrow in Fig. 6. Since there is no common point between the shaded area and the area indicated by arrow, we say that the obtained solution is the optimum solution. Hence the optimum solution to the given LP problem is: $x = 20, y = 15$ and Min $Z = 120$.

Illustration 4.8:

Solve the following LP problem using graphical method:

Maximize $\quad Z = 3x + 2y$

Subject to $\quad x - y \geq 1$

$x + y \geq 3$

$x, y \geq 0$

Solution:

Plot each constraint on graph paper by first considering it as a linear equation. Then use inequality of each constraint and find the common area where all constraints are satisfied. This area is a feasible region which is shown in Fig. 7 as shaded area.

Note that the feasible region is unbounded. The coordinates of vertices of the feasible region are: A = (0, 3) and B = (2, 1). Objective function value at A is 6 and at B is 8.

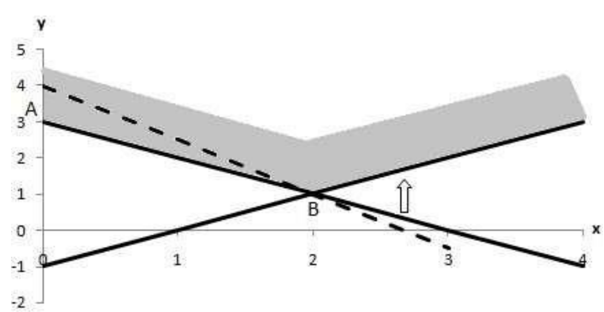


Figure 7

Since objective function Z is to be maximized, the maximum value of Z is 8 at $x = 2$ and $y = 1$.

From Fig. 7 it is clear that the feasible region is unbounded. So it may or may not be the optimal solution of the given problem. Let us plot a straight line representing the objective function $3x + 2y = 8$. This line is shown as dashed line in Fig.7. The area where the inequality $3x + 2y > 8$ satisfied is indicated by arrow in Fig. 7. Since there are infinitely many common point between the shaded area and the area indicated by arrow, we say that optimum solution does not exist or the given LP problem has an unbounded solution.

Illustration 4.9:

Solve the following LP problem using graphical method:

Maximize $Z = 3x + 2y$

Subject to $4x + 3y \leq 8$

$x + y \geq 6$

$x, y \geq 0$

Solution:

The constraints are plotted on a graph paper by first considering it as a linear equation. Then use inequality of each constraint and find the feasible region as shown shaded area in Fig. 8.

Since there is no unique point (x, y) in the shaded regions that can satisfy all the constraints simultaneously, the given LP problem has an infeasible solution.

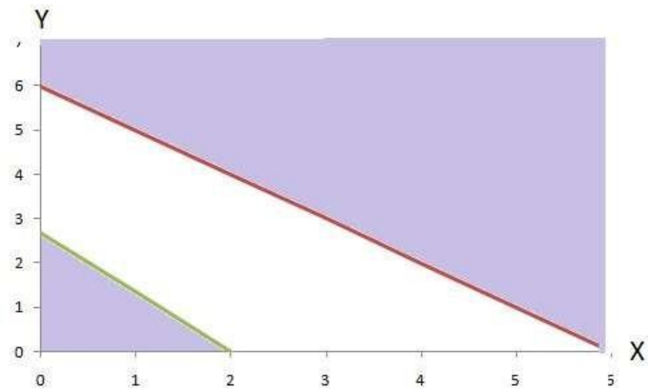


Figure 8

4.5 SIMPLEX METHOD

Simplex Method is used to obtain the optimum solution of a LP problem with two or more decision variables. Here we have discussed this method for maximization case of LP problem with two or more decision variables.

4.5.1 Steps for Solution of Linear Programming Problem by Simplex Method (Maximization Case)

Step 1: Formulation of the mathematical model.

Formulate the mathematical model of the given linear programming problem.

Check whether all the b_i $(i = 1, 2, \ldots \ldots, m)$ values are positive. If any of them is negative, then multiply the corresponding constraint by -1 to make $b_i > 0$. In doing so, change a $\leq$ type constraint to a $\geq$ type constraint, and vice-versa.

Step 2: Set up the initial solution.

Construct the initial simplex table as shown below:

Table 4.4: Initial Simplex Table

		$c_j \rightarrow$	c_1	c_2	...	c_n	0	0	...	0
Coefficient of Basic variables (c_B)	Variables in Basis B	Value of Basic Variables $b\ (= x_B)$	Variables							
			x_1	x_2	...	x_n	s_1	s_2	...	s_n
c_{B1}	s_1	$x_{B1} = b_1$	a_{11}	a_{12}	...	a_{1n}	1	0	...	0
c_{B2}	s_2	$x_{B2} = b_2$	a_{21}	a_{22}	...	a_{2n}	0	1	...	0
:	:	:	:	:		:	:	:		:
c_{Bm}	s_m	$x_{Bm} = b_m$	a_{m1}	a_{m2}	...	a_{mn}	0	0	...	1
$Z = \sum_{j=1}^{m} c_{Bj} x_{Bj}$		$z_j = \sum_{i=1}^{m} c_{Bi} x_j$	0	0	...	0	0	0	...	0
		$c_j - z_j \rightarrow$								

Step 3: Test for optimality.

Examine the values of $c_j - z_j$.

(i) If all $c_j - z_j \leq 0$, then the basic feasible solution is optimal.

(ii) If at least one $c_k - z_k > 0$ and all elements in that column are negative $(a_{ik} < 0)$, then there exists an unbounded solution to the given problem.

(iii) If at least one $c_j - z_j > 0$ and each of these has at least one positive element (i.e. a_{ij}) for some row, then the improvement in the value of objective function Z is possible.

Step 4: Select the variable to enter the basis.

If case (iii) of Step 3 holds, then select a variable that has the largest $c_j - z_j$ value to enter into the new solution. This column which is to be entered is called the key column. If the same largest value of $c_j - z_j$ is found for more than one variable, choose any one of them as entering variable.

Step 5: Select the variable to leave the basis.

Divide each number in x_B –column (i.e. b_i values) by the corresponding (but positive) number in the key column and select the row for which this ratio is non-negative and minimum. The row selected in this manner is called key row and the element intersecting key row and key column is called key element.

Step 6: Finding the new solution.

(i) Divide each element in the key row (including elements in x_B-column) by the key element, to find the new values for that row.

(ii) Perform the following elementary row operations on all rows other than key row.

Number in new row = Number in old row – [(Number above or below key element) $\times$

(Corresponding number in the new key row)]

Prepare a new table with these updated entries in all rows.

Step 7: Repeat the procedure.

Go to Step 3 and repeat the procedure until all entries in the $c_j - z_j$ row are either negative or zero.

4.5.2 Problems on Solution of Linear Programming Problem using Simplex Method

Illustration 4.10:

Use Simplex method to solve the following LPP:

Maximize $\quad Z = 4x + 3y$

Subject to $\quad 2x + y \leq 30$

$$x + 2y \leq 40$$

$$x + y \leq 25$$

$$x, y \geq 0$$

Solution:

Converting inequalities into equations by introducing slack variables s_1, s_2 and s_3 in the constraints, we get

$$2x + y + s_1 + 0s_2 + 0s_3 = 30$$

$$x + 2y + 0s_1 + s_2 + 0s_3 = 40$$

$$x + y + 0s_1 + 0s_2 + s_3 = 25$$

and the objective function becomes

$$Z = 4x + 3y + 0s_1 + 0s_2 + 0s_3$$

Initial solution is obtained by taking $x = 0$ and $y = 0$. So initially slack variables are basic variables and value of slack variables s_1, s_2 and s_3 are obtained as 30, 40 and 25 respectively.

Table 4.5: Initial Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j					Minimum Ratio= $\frac{x_B}{x}$
			4	3	0	0	0	
			x	y	s_1	s_2	s_3	
s_1	0	30	2	1	1	0	0	$\frac{30}{2} = 15$
s_2	0	40	1	2	0	1	0	$\frac{40}{1} = 40$
s_3	0	25	1	1	0	0	1	$\frac{25}{1} = 25$
$Z = 0$	$z_j = \sum C_B x_j$		0	0	0	0	0	
	$c_j - z_j$		**4**	3	0	0	0	

Note that x_j denote the j^th variable from the variables x, y, s_1, s_2 and s_3. While c_j denote a coefficient of j^th variable, C_B is a coefficient of base variable in the objective function.

Calculation of the objective function:

Here $x = 0$ and $y = 0$.

So $Z = 4x + 3y$

$= 4(0) + 3(0) = 0.$

In the initial solution not all $c_j - z_j \leq 0$. Hence the solution is not optimal and we have to modify the solution.

Since $c_j - z_j = 4$ is the largest positive value for the x – column, the x – column is the key column and the variable x will enter into the basis. The variable that is to leave the basis is determined by dividing the value in the x_B – column by the corresponding elements in the key column as shown in the Table 4.5. Since the ratio $\frac{30}{2}$ is minimum, the basic variable s_1 will leave the basis.

We calculate elements of the three rows in the new solution as below:

New Row – 1 : Old Row – 1 / Corresponding element in key column

$$: \frac{30}{2}, \frac{2}{2}, \frac{1}{2}, \frac{1}{2}, \frac{0}{2}, \frac{0}{2}$$

$$: 15, 1, \frac{1}{2}, \frac{1}{2}, 0, 0$$

New Row – 2 : (Old Row – 2) – (Corresponding element in key column) × (New Row – 1)

$$: (40, 1, 2, 0, 1, 0) - 1 \times (15, 1, \frac{1}{2}, \frac{1}{2}, 0, 0)$$

$$: 40 - 15, 1 - 1, 2 - \frac{1}{2}, 0 - \frac{1}{2}, 1 - 0, 0 - 0$$

$$: 25, 0, \frac{3}{2}, -\frac{1}{2}, 1, 0$$

New Row – 3 : (Old Row – 3) – (Corresponding element in key column) × (New Row – 1)

$$: (25, 1, 1, 0, 0, 1) - 1 \times (15, 1, \frac{1}{2}, \frac{1}{2}, 0, 0)$$

$$: 25 - 15, 1 - 1, 1 - \frac{1}{2}, 0 - \frac{1}{2}, 0 - 0, 1 - 0$$

$$: 10, 0, \frac{1}{2}, -\frac{1}{2}, 0, 1$$

The new improved solution to the given LPP is shown in Table 4.6.

Table 4.6: Improved Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j					Minimum Ratio= $\frac{x_B}{y}$
			4	3	0	0	0	
			x	y	s_1	s_2	s_3	
x	4	15	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	30
s_2	0	25	0	$\frac{3}{2}$	$-\frac{1}{2}$	1	0	$\frac{50}{3}$
s_3	0	10	0	$\frac{1}{2}$	$-\frac{1}{2}$	0	1	20
$Z = 60$	$z_j = \sum C_B x_j$		4	2	2	0	0	
	$c_j - z_j$		0	**1**	−2	0	0	

Calculation of the objective function:

Here $x = 15$ and $y = 0$. So $Z = 4x + 3y = 4(15) + 3(0) = 60$.

Since not all $c_j - z_j \leq 0$, the solution is not optimal and we have to modify the solution.

Here $c_j - z_j = 1$ is the largest positive value for the y – column, the y – column is the key column and the variable y will enter into the basis. It is also seen from the Table 4.6 that the basic variable s_2 will leave the basis.

We calculate elements of all the three rows in the new solution as below:

New Row – 2 : Old Row – 2 / Corresponding element in key column

$: (25, 0, \frac{3}{2}, -\frac{1}{2}, 1, 0) \div (\frac{3}{2})$

$: \frac{50}{3}, 0, 1, -\frac{1}{3}, \frac{2}{3}, 0$

New Row – 1 : (Old Row – 1) – (Corresponding element in key column) × (New Row – 2)

$: (15, 1, \frac{1}{2}, \frac{1}{2}, 0, 0) - (\frac{1}{2}) \times (\frac{50}{3}, 0, 1, -\frac{1}{3}, \frac{2}{3}, 0)$

$: 15 - \frac{25}{3}, 1 - 0, \frac{1}{2} - \frac{1}{2}, \frac{1}{2} + \frac{1}{6}, 0 - \frac{1}{3}, 0 - 0$

$: \frac{20}{3}, 1, 0, \frac{2}{3}, -\frac{1}{3}, 0$

New Row – 3 : (Old Row – 3) – (Corresponding element in key column) × (New Row – 2)

$: (10, 0, \frac{1}{2}, -\frac{1}{2}, 0, 1) - (\frac{1}{2}) \times (\frac{50}{3}, 0, 1, -\frac{1}{3}, \frac{2}{3}, 0)$

$: 10 - \frac{50}{6}, 0 - 0, \frac{1}{2} - \frac{1}{2}, -\frac{1}{2} + \frac{1}{6}, 0 - \frac{1}{3}, 1 - 0$

$: \frac{5}{3}, 0, 0, -\frac{1}{3}, -\frac{1}{3}, 1$

The new improved solution to the given LPP is shown in Table 4.7.

Table 4.7: Improved Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j				
			4	3	0	0	0
			x	y	s_1	s_2	s_3
x	4	$\frac{20}{3}$	1	0	$\frac{2}{3}$	$-\frac{1}{3}$	0
y	3	$\frac{50}{3}$	0	1	$-\frac{1}{3}$	$\frac{2}{3}$	0
s_3	0	$\frac{5}{3}$	0	0	$-\frac{1}{3}$	$-\frac{1}{3}$	1
$Z = 70$	$z_j = \sum C_B x_j$		4	3	$\frac{5}{3}$	$\frac{2}{3}$	0
	$c_j - z_j$		0	0	$-\frac{5}{3}$	$-\frac{2}{3}$	0

Calculation of the objective function:

Here $x = \frac{20}{3}$ and $y = \frac{50}{3}$. So $Z = 4x + 3y = 4\left(\frac{20}{3}\right) + 3\left(\frac{50}{3}\right) = \frac{230}{3}$.

It is noted that in the last iteration, all $c_j - z_j \leq 0$. Hence the solution is optimum and the solution is given by Maximum $Z = \frac{230}{3}$ for $x = \frac{20}{3}$ and $y = \frac{50}{3}$.

Illustration 4.11:

Solve the following LPP using Simplex method:

Maximize $\quad Z = x_1 - x_2 + x_3$

Subject to

$$2x_1 + x_2 - 3x_3 \leq 40$$

$$x_1 \quad + x_3 \leq 25$$

$$2x_2 + 3x_3 \leq 32$$

$$x_1, x_2, x_3 \geq 0$$

Solution:

Converting inequalities into equations by introducing slack variables s_1, s_2 and s_3 in the constraints, we get

$$2x_1 + x_2 - 3x_3 + s_1 + 0s_2 + 0s_3 = 40$$

$$x_1 \quad + x_3 + 0s_1 + s_2 + 0s_3 = 25$$

$$2x_2 + 3x_3 + 0s_1 + 0s_2 + s_3 = 32$$

and the objective function becomes

$$Z = x_1 - x_2 + x_3 + 0s_1 + 0s_2 + 0s_3$$

Initial solution is obtained by setting $x_1 = 0$, $x_2 = 0$ and $x_3 = 0$. So initially slack variables are basic variables and value of slack variables s_1, s_2 and s_3 are obtained as 40, 25 and 32 respectively

Table 4.8: Initial Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j						Minimum Ratio= $\frac{x_B}{x_3}$
			1	–1	1	0	0	0	
			x_1	x_2	x_3	s_1	s_2	s_3	
s_1	0	40	2	1	–3	1	0	0	---
s_2	0	25	1	0	1	0	1	0	$\frac{25}{1}$
s_3	0	32	0	2	3	0	0	1	$\frac{32}{3}$
$Z = 0$	$z_j = \sum C_B x_j$		0	0	0	0	0	0	
	$c_j - z_j$		1	–1	**1**	0	0	0	

Here x_j denote the jth variable from the variables x_1, x_2, x_3, s_1, s_2 and s_3.

Calculation of the objective function:

Here $x_1 = 0, x_2 = 0$ and $x_3 = 0$. So $Z = x_1 - x_2 + x_3 = 0 - 0 + 0 = 0$.

In the initial solution not all $c_j - z_j \leq 0$. Hence the solution is not optimal and we have to modify the solution.

Since $c_j - z_j = 1$ is the largest positive value for the x_1 – column and x_3 – column, we can choose any one of these two as key column. Let us choose the x_3 – column as the key column and the variable x_3 will enter into the basis. The variable that is to leave the basis is determined by dividing the value in the x_B – column by the corresponding elements (only positive) in the key column as shown in the Table 4.8. Since the ratio 32/3 is minimum, the basic variable s_3 will leave the basis. Note that division by negative key element is not permissible in finding minimum ratio.

We calculate elements of the three rows in the new solution as below: New Row – 3: Old Row – 3 / Corresponding element in key column

$$: \frac{32}{3} \; \frac{0}{3} \; \frac{2}{3} \; \frac{3}{3} \; \frac{0}{3} \; \frac{0}{3} \; \frac{1}{3}$$

$$: \frac{32}{3}, 0, \frac{2}{3}, 1,0,0, \frac{1}{3}$$

New Row – 1 : (Old Row – 1) – (Corresponding element in key column) × (New Row – 3)

$: (40, 2, 1, -3, 1, 0, 0) - (-3) \times (\frac{32}{3}, 0, \frac{2}{3}, 1, 0, 0, \frac{1}{3})$

$: 40 + 32, 2 + 0, 1 + 2, -3 + 3, 1 + 0, 0 + 0, 0 + 1$

$: 72, 2, 3, 0, 1, 0, 1$

New Row – 2 : (Old Row – 2) – (Corresponding element in key column) × (New Row – 3)

$: (25, 1, 0, 1, 0, 1, 0) - 1 \times (\frac{32}{3}, 0, \frac{2}{3}, 1, 0, 0, \frac{1}{3})$

$: 25 - \frac{32}{3}, 1 - 0, 0 - \frac{2}{3}, 1 - 1, 0 - 0, 1 - 0, 0 - \frac{1}{3}$

$: \frac{43}{3}, 1, -\frac{2}{3}, 0, 0, 1, -\frac{1}{3}$

The new improved solution to the given LPP is shown in Table 4.9.

Table 4.9: Improved Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j						Minimum Ratio=$\frac{x_B}{x_2}$
			1	-1	1	0	0	0	
			x_1	x_2	x_3	s_1	s_2	s_3	
s_1	0	72	2	3	0	1	0	1	$\frac{72}{2}$
s_2	0	$\frac{43}{3}$	1	$-\frac{2}{3}$	0	0	1	$-\frac{1}{3}$	$\frac{43}{3}$
x_3	1	$\frac{32}{3}$	0	$\frac{2}{3}$	1	0	0	$\frac{1}{3}$	---
$Z = \frac{32}{3}$	$z_j = \sum C_B x_j$		0	$\frac{2}{3}$	1	0	0	$\frac{1}{3}$	
	$c_j - z_j$		**1**	$-\frac{5}{3}$	0	0	0	$-\frac{1}{3}$	

Calculation of the objective function:

Here $x_1 = 0, x_2 = 0$ and $x_3 = \frac{32}{3}$. So $Z = x_1 - x_2 + x_3 = 0 - 0 + \frac{32}{3} = \frac{32}{3}$.

Since not all $c_j - z_j \leq 0$, the solution is not optimal and we have to modify the solution.

Here $c_j - z_j = 1$ is the largest positive value for the x_1 – column, the x_1 – column is the key column and the variable x_1 will enter into the basis. It is also seen from the Table 4.9 that the basic variable s_2 will leave the basis.

We calculate elements of all the three rows in the new solution as below:

New Row – 2 : Old Row – 2 / Corresponding element in key column

$: (\frac{43}{3}, 1, -\frac{2}{3}, 0, 0, 1, -\frac{1}{3}) \div 1$

$: \frac{43}{3}, 1, -\frac{2}{3}, 0, 0, 1, -\frac{1}{3}$

New Row – 1 : (Old Row – 1) – (Corresponding element in key column) × (New Row – 2)

$$: (72, 2, 3, 0, 1, 0, 1) - 2 \times \left(\frac{43}{3}, 1, -\frac{2}{3}, 0, 0, 1, -\frac{1}{3}\right)$$

$$: 72 - \frac{86}{3}, 2 - 2, 3 + \frac{4}{3}, 0 - 0, 1 - 0, 0 - 2, 1 + \frac{2}{3}$$

$$: \frac{130}{3}, 0, \frac{13}{3}, 0, 1, -2\ \frac{5}{3}$$

New Row – 3 : (Old Row – 3) – (Corresponding element in key column) × (New Row – 2)

$$: \left(\frac{32}{3}, 0, \frac{2}{3}, 1, 0, 0, \frac{1}{3}\right) - 0 \times \left(\frac{43}{3}, 1, -\frac{2}{3}, 0, 0, 1, -\frac{1}{3}\right)$$

$$: \frac{32}{3}, 0, \frac{2}{3}, 1, 0, 0, \frac{1}{3}$$

The new improved solution to the given LPP is shown in Table 4.10.

Table 4.10: Improved Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j					
			1	−1	1	0	0	0
			x_1	x_2	x_3	s_1	s_2	s_3
s_1	0	$\frac{130}{3}$	0	$\frac{13}{3}$	0	1	−2	$\frac{5}{3}$
x_1	1	$\frac{43}{3}$	1	$-\frac{2}{3}$	0	0	1	$-\frac{1}{3}$
x_3	1	$\frac{32}{3}$	0	$\frac{2}{3}$	1	0	0	$\frac{1}{3}$
$Z = 25$	$z_j = \sum C_B x_j$		1	0	1	0	1	0
	$c_j - z_j$		0	−1	0	0	−1	0

Calculation of the objective function:

Here $x_1 = \frac{43}{3}$, $x_2 = 0$ and $x_3 = \frac{32}{3}$. So $Z = x_1 - x_2 + x_3 = \frac{43}{3} - 0 + \frac{32}{3} = \frac{75}{3} = 25$.

Since all $c_j - z_j \leq 0$ in the last iteration, the solution is optimum and the solution is given by Maximum $Z = 25$ for $x_1 = \frac{43}{3}$, $x_2 = 0$ and $x_3 = \frac{32}{3}$. Note that slack variable s_1 in basis has a value $\frac{130}{3}$ in the final iteration but it will not affect in calculating Z as the coefficient of s_1 is zero in the objective function.

Illustration 4.12:

Solve the following LPP using Simplex method:

Maximize $\quad Z = 3x_1 + 5x_2 + 4x_3$

Subject to $\quad 2x_1 + 3x_2 \leq 8$

$$2x_2 + 5x_3 \leq 10$$

$$3x_1 + 2x_2 + 4x_3 \leq 15$$

$$x_1, x_2, x_3 \geq 0$$

Solution:

Converting inequalities into equations by introducing slack variables s_1, s_2 and s_3 in the constraints, we get

$$x_1 + 3x_2 + 0x_3 + s_1 + 0s_2 + 0s_3 = 8$$

$$0x_1 + 2x_2 + 5x_3 + 0s_1 + s_2 + 0s_3 = 10$$

$$3x_1 + 2x_2 + 4x_3 + 0s_1 + 0s_2 + s_3 = 15$$

and the objective function becomes

$$Z = 3x_1 + 5x_2 + 4x_3 + 0s_1 + 0s_2 + 0s_3$$

Initial solution is obtained in Table 4.11 by setting $x_1 = 0$, $x_2 = 0$ and $x_3 = 0$. Initially slack variables are basic variables and value of slack variables s_1, s_2 and s_3 are obtained as 8, 10 and 15 respectively.

Table 4.11: Initial Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j						Minimum Ratio= $\frac{x_B}{x_2}$
			3	5	4	0	0	0	
			x_1	x_2	x_3	s_1	s_2	s_3	
s_1	0	8	2	3	0	1	0	0	$\frac{8}{3}$
s_2	0	10	0	2	5	0	1	0	$\frac{10}{2}$
s_3	0	15	3	2	4	0	0	1	$\frac{15}{2}$
$Z = 0$	$z_j = \sum C_B x_j$		0	0	0	0	0	0	
	$c_j - z_j$		3	**5**	4	0	0	0	

Here x_j denote the j[th] variable from the variables x_1, x_2, s_1, s_2 and s_3.

Calculation of the objective function:

Here $x_1 = 0, x_2 = 0$ and $x_3 = 0$. So $Z = 3x_1 + 5x_2 + 4x_3 = 3(0) + 5(0) + 4(0) = 0$.

In the initial solution not all $c_j - z_j \leq 0$. Hence the solution is not optimal and we have to modify the solution.

Since $c_j - z_j = 5$ is the largest positive value for the x_2 – column, the x_2 – column is the key column and the variable x_2 will enter into the basis. The variable that is to leave the basis is determined by dividing the value in the x_B – column by the corresponding elements in the key column as shown in the Table 4.11. Since the ratio is minimum, the basic variable s_1 will leave the basis.

We calculate elements of the three rows in the new solution as below:

New Row – 1 : Old Row – 1 / Corresponding element in key column

$$: \frac{8}{3}, \frac{2}{3}, \frac{3}{3}, \frac{0}{3}, \frac{1}{3}, \frac{0}{3}, \frac{0}{3}$$

$$: \frac{8}{3}, \frac{2}{3}, 1, 0, \frac{1}{3}, 0, 0$$

New Row – 2 : (Old Row – 2) – (Corresponding element in key column) × (New Row – 1)

$$: (10, 0, 2, 5, 0, 1, 0) - 2 \times \left(\frac{8}{3}, \frac{2}{3}, 1, 0, \frac{1}{3}, 0, 0\right)$$

$$: 10 - \frac{16}{3}, 0 - \frac{4}{3}, 2 - 2, 5 - 0, 0 - \frac{2}{3}, 1 - 0, 0 - 0$$

$$: \frac{14}{3}, -\frac{4}{3}, 0, 5, -\frac{2}{3}, 1, 0$$

New Row – 3 : (Old Row – 3) – (Corresponding element in key column) × (New Row – 1)

$$: (15, 3, 2, 4, 0, 0, 1) - 2 \times \left(\frac{8}{3}, \frac{2}{3}, 1, 0, \frac{1}{3}, 0, 0\right)$$

$$: 15 - \frac{16}{3}, 3 - \frac{4}{3}, 2 - 2, 4 - 0, 0 - \frac{2}{3}, 0 - 0, 1 - 0$$

$$: \frac{29}{3}, \frac{5}{3}, 0, 4, -\frac{2}{3}, 0, 1$$

The new improved solution to the given LPP is shown in Table 4.12.

Table 4.12: Improved Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j						Minimum Ratio=$\frac{x_B}{x_3}$
			3	5	4	0	0	0	
			x_1	x_2	x_3	s_1	s_2	s_3	
x_2	5	$\frac{8}{3}$	$\frac{2}{3}$	1	0	$\frac{1}{3}$	0	0	---
s_2	0	$\frac{14}{3}$	$-\frac{4}{3}$	0	5	$-\frac{2}{3}$	1	0	$\frac{14}{15}$
s_3	0	$\frac{29}{3}$	$\frac{5}{3}$	0	4	$-\frac{2}{3}$	0	1	$\frac{29}{12}$
$Z=\frac{40}{3}$	$z_j=\sum C_B x_j$		$\frac{10}{3}$	5	0	$\frac{5}{3}$	0	0	
	$c_j - z_j$		$-\frac{1}{3}$	0	**4**	$-\frac{5}{3}$	0	0	

Calculation of the objective function:

Here $x_1 = 0, x_2 = \frac{8}{3}$ and $x_3 = 0$. So $Z = 3x_1 + 5x_2 + 4x_3 = 3(0) + 5(\frac{8}{3}) + 4(0) = \frac{40}{3}$.

Since not all $c_j - z_j \leq 0$, the solution is not optimal and we have to modify the solution.

Here $c_j - z_j = 4$ is the largest positive value for the x_3 – column, the x_3 – column is the key column and the variable x_3 will enter into the basis. It is also seen from the Table 4.12 that the basic variable s_2 will leave the basis. Note that division by zero is not permissible in finding minimum ratio.

We calculate elements of all the three rows in the new solution as below:

New Row – 2 : Old Row – 2 / Corresponding element in key column

$: (\frac{14}{3}.-\frac{4}{3}, 0, 5, -\frac{2}{3}, 1, 0) \div 5$

$: \frac{14}{15}, -\frac{4}{15}, 0, 1, -\frac{2}{15}, \frac{1}{5}, 0$

New Row – 1 : (Old Row – 1) – (Corresponding element in key column) × (New Row – 2)

$: (\frac{8}{3}, \frac{2}{3}, 1, 0, \frac{1}{3}, 0, 0) - 0 \times (\frac{14}{15}, -\frac{4}{15}, 0, 1, -\frac{2}{15}, \frac{1}{5}, 0)$

$: \frac{8}{3}, \frac{2}{3}, 1, 0, \frac{1}{3}, 0, 0$

New Row – 3 : (Old Row – 3) – (Corresponding element in key column) × (New Row – 2)

$: (\frac{29}{3}, \frac{5}{3}, 0, 4, -\frac{2}{3}, 0, 1) - 4 \times (\frac{14}{15}, -\frac{4}{15}, 0, 1, -\frac{2}{15}, \frac{1}{5}, 0)$

$: \frac{29}{3} - \frac{56}{15}, \frac{5}{3} + \frac{16}{15}, 0 - 0, 4 - 4, -\frac{2}{3} + \frac{8}{15}, 0 - \frac{4}{5}, 1 - 0$

$: \frac{89}{15}, \frac{41}{15}, 0, 0, -\frac{2}{15}, -\frac{4}{5}, 1$

The new improved solution to the given LPP is shown in Table 4.13.

Table 4.13: Improved Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j						Minimum Ratio=$\frac{x_B}{x_1}$
			3	5	4	0	0	0	
			x_1	x_2	x_3	s_1	s_2	s_3	
x_2	5	$\frac{8}{3}$	$\frac{2}{3}$	1	0	$\frac{1}{3}$	0	0	$\frac{8}{2}$
x_3	4	$\frac{14}{15}$	$-\frac{4}{15}$	0	1	$-\frac{2}{15}$	$\frac{1}{5}$	0	---
s_3	0	$\frac{89}{15}$	$\frac{41}{15}$	0	0	$-\frac{2}{15}$	$-\frac{4}{5}$	1	$\frac{89}{41}$
$Z = \frac{256}{15}$	$z_j = \sum C_B x_j$		$\frac{34}{15}$	5	4	$\frac{17}{15}$	$\frac{4}{5}$	0	
	$c_j - z_j$		$\frac{11}{15}$	0	0	$-\frac{17}{15}$	$-\frac{4}{5}$	0	

Calculation of the objective function:

Here $x_1 = 0, x_2 = \frac{8}{3}$ and $x_3 = \frac{14}{15}$.

So $Z = 3x_1 + 4x_2 + 5x_3 = 3(0) + 5\left(\frac{8}{3}\right) + 4\left(\frac{14}{15}\right) = \frac{40}{3} + \frac{56}{15} = \frac{256}{15}$

Since not all $c_j - z_j \leq 0$, the solution is not optimal and we have to modify the solution.

Here $c_j - z_j = \frac{11}{15}$ is the largest positive value for the x_1 – column, the x_1 – column is the key column and the variable x_1 will enter into the basis. It is also seen from the Table 4.13 that the basic variable s_3 will leave the basis. Note that division by negative key element is not permissible in finding minimum ratio.

We calculate elements of all the three rows in the new solution as below:

New Row -3 : Old Row – 3 / Corresponding element in key column

$$:\left(\frac{89}{15}, \frac{41}{15}, 0,0, -\frac{2}{15}, -\frac{4}{5}, 1\right) \div \left(\frac{41}{15}\right)$$

$$:\frac{89}{41}, 1,0,0, -\frac{2}{41}. \frac{12}{41}. \frac{15}{41}$$

New Row – 1 : (Old Row – 1) – (Corresponding element in key column) × (New Row – 3)

$$:\left(\frac{8}{3}, \frac{2}{3}, 1, 0, \frac{1}{3}, 0, 0\right) - \left(\frac{2}{3}\right) \times \left(\frac{89}{41}, 1, 0, 0, -\frac{2}{41}. -\frac{12}{41}. \frac{15}{41}\right)$$

$$\left(\frac{8}{3},\frac{2}{3},1,0,\frac{1}{3},0,0\right)-\left(\frac{178}{123},\frac{2}{3},0,0\frac{4}{123},\frac{24}{123},\frac{30}{123}\right)$$

$$:\frac{150}{123},0,1,0,\frac{45}{123}.\frac{24}{123}.-\frac{30}{123}$$

$$:\frac{50}{41},0,1,0,\frac{15}{41},\frac{8}{41}.-\frac{10}{41}$$

New Row – 2 : (Old Row – 2) – (Corresponding element in key column) × (New Row – 3)

$$:(\frac{14}{15},-\frac{4}{15},0,1,-\frac{2}{15},\frac{1}{5},0)-(-\frac{4}{15})\times(\frac{89}{41},1,0,0,-\frac{2}{41}.-\frac{12}{41}.\frac{15}{41})$$

$$:(\frac{14}{15},-\frac{4}{15},0,1,-\frac{2}{15},\frac{1}{5},0)+(\frac{356}{615}.\frac{4}{15},0,0,-\frac{8}{615}.-\frac{48}{615},\frac{4}{41})$$

$$:\frac{930}{615},0,0,1,-\frac{90}{615}.\frac{75}{615}.\frac{4}{41}$$

$$:\frac{62}{41},0,0,1,-\frac{6}{41}.\frac{5}{41}.\frac{4}{41}$$

The new improved solution to the given LPP is shown in Table 4.14.

Table 4.14: Improved Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j					
			3	5	4	0	0	0
			x_1	x_2	x_3	s_1	s_2	s_3
x_2	5	$\frac{50}{41}$	0	1	0	$\frac{15}{41}$	$\frac{8}{41}$	$\frac{10}{41}$
x_3	4	$\frac{62}{41}$	0	0	1	$\frac{6}{41}$	$\frac{5}{41}$	$\frac{4}{41}$
x_1	3	$\frac{89}{41}$	1	0	0	$\frac{2}{41}$	$\frac{12}{41}$	$\frac{15}{41}$
$Z = \frac{765}{41}$	$z_j = \sum C_B x_j$		3	5	4	$\frac{45}{41}$	$\frac{24}{41}$	$\frac{11}{41}$
	$c_j - z_j$		0	0	0	$-\frac{45}{41}$	$-\frac{24}{41}$	$-\frac{11}{41}$

Calculation of the objective function:

Here $x_1 = \frac{89}{41}, x_2 = \frac{50}{41}$ and $x_3 = \frac{62}{41}$.

$$\text{So Z } = 3x_1 + 5x_2 + 4x_3 = 3\left(\frac{89}{41}\right) + 5\left(\frac{50}{41}\right) + 4\left(\frac{62}{41}\right) = \frac{267}{41} + \frac{250}{41} = \frac{248}{41} = \frac{765}{41}$$

Since all $c_j - z_j \leq 0$ in the last iteration, the solution is optimum and the solution is given by Maximum $Z = \frac{765}{41}$ for $x_1 = \frac{89}{41}$, $x_2 = \frac{50}{41}$ and $x_3 = \frac{62}{41}$.

Illustration 4.13:

Solve the following LPP using Simplex method:

Maximize $\quad Z = 2x_1 + 3x_2$

Subject to $\quad -x_1 + 2x_2 \leq 4$

$$x_1 + x_2 \leq 6$$

$$x_1 + 3x_2 \leq 9$$

x_1, x_2 are unrestricted.

Solution:

Since both the decision variable x_1 and x_2 are unrestricted, we introduce non negative variables x'_1, x''_1, x'_2, x''_2 such that $x_1 = x'_1 - x''_1$ and $x_2 = x'_2 - x''_2$.

The given LPP can be written as

Maximize $\quad Z = 2x'_1 - 2x''_1 + 3x'_2 - 3x''_2$

Subject to $\quad -x'_1 + x''_1 + 2x'_2 - 2x''_2 \leq 4$

$$x'_1 - x''_1 + x'_2 - x''_2 \leq 6$$

$$x'_1 - x''_1 + 3x'_2 - 3x''_2 \leq 9$$

$$x'_1, x''_1, x'_2, x''_2 \geq 0.$$

Converting inequalities into equations by introducing slack variables s_1, s_2 and s_3 in the constraints, we get

$$-x'_1 + x''_1 + 2x'_2 - 2x''_2 + s_1 + 0s_2 + 0s_3 = 4$$

$$x'_1 - x''_1 + x'_2 - x''_2 + 0s_1 + s_2 + 0s_3 = 6$$

$$x'_1 - x''_1 + 3x'_2 - 3x''_2 + 0s_1 + 0s_2 + s_3 = 9$$

$$x'_1, x''_1, x'_2, x''_2, s_1, s_2, s_3 \geq 0.$$

and the objective function becomes

$$Z = 2x'_1 - 2x''_1 + 3x'_2 - 3x''_2 + 0s_1 + 0s_2 + 0s_3$$

Initial solution is obtained by setting $x'_1 = 0$, $x''_1 = 0$, $x'_2 = 0$ and $x''_2 = 0$. So initially slack variables are basic variables and value of slack variables s_1, s_2 and s_3 are obtained as 4, 6 and 9 respectively.

Table 4.15: Initial Solution

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j: 2	−2	3	−3	0	0	0	Minimum Ratio$=\frac{x_B}{x'_2}$
			x'_1	x''_1	x'_2	x''_2	s_1	s_2	s_3	
s_1	0	4	−1	1	2	−2	1	0	0	$\frac{4}{2}$
s_2	0	6	1	−1	1	−1	0	1	0	$\frac{6}{1}$
s_3	0	9	1	−1	3	−3	0	0	1	$\frac{9}{3}$
$Z = 0$	$z_j = \sum C_B x_j$		0	0	0	0	0	0	0	
	$c_j - z_j$		2	−2	3	−3	0	0	0	

In the initial solution not all $c_j - z_j \leq 0$. Hence the solution is not optimal. So we have to modify the solution.

From Table 4.15 it is clear that in the modified simplex table, x'_2 will enter into the basis and s_1 will exit from the basis.

Following usual calculations explained in the previous illustrations of simplex method, we can construct the successive modified simplex tables as below:

Table 4.16: Modified Simplex Table

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j: 2	−2	3	−3	0	0	0	Minimum Ratio$=\frac{x_B}{x'_1}$
			x'_1	x''_1	x'_2	x''_2	s_1	s_2	s_3	
x'_2	3	2	$-\frac{1}{2}$	$\frac{1}{2}$	1	−1	$\frac{1}{2}$	0	0	---
s_2	0	4	$\frac{3}{2}$	$-\frac{3}{2}$	0	0	$-\frac{1}{2}$	1	0	$\frac{8}{3}$
s_3	0	3	$\frac{5}{2}$	$-\frac{5}{2}$	0	0	$-\frac{3}{2}$	0	1	$\frac{6}{5}$
$Z = 6$	$z_j = \sum C_B x_j$		$-\frac{3}{2}$	$\frac{3}{2}$	3	−3	$\frac{3}{2}$	0	0	
	$c_j - z_j$		$\frac{7}{2}$	$-\frac{7}{2}$	0	0	$-\frac{3}{2}$	0	0	

From Table 4.16 it is clear that in the modified simplex table, x' will enter into the basis and s_3 will exit from the basis.

From Table 4.17 it is clear that in the modified simplex table, s_1 will enter into the basis and s_2 will exit from the basis.

Table 4.17: Modified Simplex Table

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j							Minimum Ratio=$\frac{x_B}{s_1}$
			2	−2	3	−3	0	0	0	
			x'_1	x''_1	x'_2	x''_2	s_1	s_2	s_3	
x'_2	3	$\frac{13}{5}$	0	0	1	−1	$\frac{1}{5}$	0	$\frac{1}{5}$	$\frac{13}{1}$
s_2	0	$\frac{11}{5}$	0	0	0	0	$\frac{2}{5}$	1	$-\frac{3}{5}$	$\frac{11}{2}$
x'_1	2	$\frac{6}{5}$	1	−1	0	0	$-\frac{3}{5}$	0	$\frac{2}{5}$	---
$Z=\frac{51}{5}$	$z_j = \sum C_B x_j$		2	−2	3	−3	$-\frac{3}{5}$	0	$\frac{7}{5}$	
	$c_j - z_j$		0	0	0	0		0	$-\frac{7}{5}$	

From Table 4.17 it is clear that in the modified simplex table, S1 will enter into the basis and S2 will exit from the basis.

Table 4.18: Modified Simplex Table

Base variable B	Coefficient of base variable C_B	Value of base variable x_B	c_j						
			2	−2	3	−3	0	0	0
			x'_1	x''_1	x'_2	x''_2	s_1	s_2	s_3
x_2	3	$\frac{3}{2}$	0	0	1	−1	0	$\frac{1}{2}$	$-\frac{1}{2}$
s_1	0	$\frac{11}{2}$	0	0	0	0	1	$\frac{5}{2}$	$-\frac{3}{2}$
x_1	2	$\frac{9}{2}$	1	−1	0	0	0	$\frac{3}{2}$	$-\frac{1}{2}$
$Z=\frac{27}{2}$	$z_j = \sum C_B x_j$		2	−2	3	−3	0	$\frac{3}{2}$	$\frac{1}{2}$
	$c_j - z_j$		0	0	0	0	0	$-\frac{3}{2}$	$-\frac{1}{2}$

Since all $c_j - z_j \leq 0$ in the last iteration, the solution is optimal. Optimum solution is $x_1' = \frac{9}{2}$, $x_1'' = 0, x_2' = \frac{3}{2}$ and $x_2'' = 0$.

Hence the solution of the original problem is

$x_1 = x_1' - x_1'' = \frac{9}{2} - 0 = \frac{9}{2}$ and $x_2 = x_2' - x_2'' = \frac{3}{2} - 0 = \frac{3}{2}$

and $Z = 2\left(\frac{9}{2}\right) + 3\left(\frac{3}{2}\right) = \frac{27}{2}$.

❖ **CHECK YOUR PROGRESS**

- **Answer the following multiple choice questions.**

Que. 1 To find solution of the LP problem, values of__________are to be determined.

(a) constraints (b) decision variables

(c) opportunity costs (d) all of these

Que. 2 Which of the following specifies the goal of solving the LP problem?

(a) objective function (b) constraints

(c) decision variables (d) opportunity costs

Que. 3 The type of constraint which specifies maximum capacity of a resource is '__________' constraint.

(a) $<$ (b) $>$

(c) $\leq$ (d) $\geq$

Que. 4 When the feasible region is such that the value of objective function can extend to infinity, it is called a case of _________.

(a) infeasible solution (b) alternate optimal

(c) unique solution (d) unbounded solution

Que. 5 The incoming variable column in the Simplex algorithm is called____________.

(a) incoming column (b) key column

(c) variable column (d) none of these

Que. 6 The intersection value of key column and key row in the Simplex algorithm is called _________.

(a) basic element (b) important element

(c) key element (d) all of these

Que. 7 A variable added to the left hand side of $\leq$ constraint to convert it into equality is called__________.

(a) surplus variable (b) additional variable

(c) artificial variable (d) slack variable

Que. 8 In Simplex algorithm, the basic feasible solution of LP problem is optimum when all $c_j - z_j$ values are__________.

(a) either zero or positive (b) either zero or negative

(c) only positive (d) only negative

- **Answer the following questions in brief.**

Que. 1 Define (i) Solution of LPP (ii) Basic solution of LPP.

Que. 2 Define unbounded solution.

Que. 3 State any four applications of linear programming.

Que. 4 Define feasible and infeasible solution of LP problem.

Que. 5 What is objective function?

Que. 6 In simplex algorithm, how you determine incoming variable?

Que. 7 Write a standard form of minimization linear programming problem.

Que. 8 Write steps for linear programming model formulation.

- **Answer the following questions in detail.**

Que. 1 Discuss on assumptions of linear programming model.

Que. 2 The ABC Company has been a producer of picture tubes for television sets and certain circuits for radios. It has built a new plant that can operate 48 hours per week. Production of one picture tube for television in the new plant will require 2 hours and production of one circuit for radio will require 3 hours. Each picture tube will contribute Rs. 40 to profits while a circuit for radio will contribute Rs. 80 to profits. The marketing departments, after extensive research, have determined that a maximum of 15 picture tubes for television and 10 circuits for radio can be sold each week. Formulate this LPP and solve it by graphical method in order to maximize the profit.

Que. 3 Write the steps for solving linear programming problem by graphical method.

Que. 4 Solve the following LP problems using graphical method.

(i) Minimize $Z = 3x + 2y$

Subject to
$$\begin{aligned} 5x + \ \ y &\geq 10 \\ x + \ \ y &\geq 6 \\ x + 4y &\geq 12 \\ x, y &\geq 0 \end{aligned}$$

(ii) Maximize $Z = 3x + 5y$

Subject to
$$\begin{aligned} x + 2y &\leq 2000 \\ 2x + y &\leq 1500 \\ y &\leq 600 \\ x, \ &\geq 0 \end{aligned}$$

Que. 5 Write the algorithm to solve linear programming problem using Simplex method for maximization case.

Que. 6 Solve the following LP problems using Simplex method.

(i) Maximize $Z = 3x_1 + 2x_2 + 5x_3$

Subject to
$$\begin{aligned} x_1 + 2x_2 + x_3 &\leq 430 \\ 3x_1 \qquad + 2x_3 &\leq 460 \\ x_1 + 4x_2 \qquad &\leq 420 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}$$

(ii) Maximize $Z = x_1 - 3x_2 + 2x_3$

Subject to
$$\begin{aligned} 3x_1 - \ \ x_2 + 3x_3 &\leq 7 \\ -2x_1 + 4x_2 \qquad &\leq 12 \\ -4x_1 + 3x_2 + 8x_3 &\leq 10 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}$$

UNIT- 5 TRANSPORTATION MODEL

5.1 INTRODUCTION

The objective of a transportation problem is to transport a commodity from various sources to different destinations at a total minimum cost. In this unit we study different methods for finding number of units to be transported from different sources to different destinations in order to minimize the total transportation cost. These methods does not give the optimal cost, hence such a solution found by these methods is called initial solution of the problem. To understand a transportation problem more clearly, refer mathematical model of transportation problem given in the next section.

5.2 IMPORTANCE

The transportation model is concerned with selecting the routes between supply and demand points in such a way that the total cost of transportation is minimized subject to the constraints of supply at sources and demand at destinations. Thus it is a special purpose problem of linear programming. Simplex method is not sufficient for this kind of problems while the transportation model gives more effective solution to such problems.

The transportation model is useful in multi-plant company to minimize the cost of transporting new materials from various centers to different manufacturing plants. It is also useful for a multi-plant-multi-market company to decide the cost of transportation of finished goods from different manufacturing plants to the different distribution centers.

5.3 MATHEMATICAL MODEL OF TRANSPORTATION PROBLEM

It involves a large number of shipping routes from several supply origins to several demand destinations. The objective is to minimize the transportation cost or time.

Table 5.1: General Transportation Table

From \ To	D_1	D_2	… …	D_4	Supply a_i
S_1	c_{11} x_{11}	c_{12} x_{12}	… …	c_{1n} x_{1n}	a_1
S_2	c_{21} x_{21}	c_{22} x_{22}	… …	c_{2n} x_{2n}	a_2
:	:	:	:	:	:
S_m	c_{m1} x_{m1}	c_{m2} x_{m2}	… …	c_{mn} x_{mn}	a_m
Demand b_j	b_1	b_2	… …	b_n	Total Supply = Total Demand

❖ **Remarks:**

1. In this problem there are $(m + n)$ constraints and x_{11}, x_{12} etc. are allocations.
2. A necessary and sufficient condition for the existence of a feasible solution to the transportation problem is

   Total Supply = Total Demand

   OR $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$ (Also called rim conditions)

5.4 INITIAL FEASIBLE SOLUTION

The optimum solution of a transportation problem can be obtained by using the initial solutionof the problem. We will discuss three different methods to obtain initial solution. Each of theseethods may produce different solutions. This initial solution is tested for its optimality; if it is not optimal the solution is modified and again checks for optimality. The process is repeated until the optimum solution reached. Following is a sequence of steps to obtain an optimum solution of any transportation problem.

Step 1: Construct a transportation table.

Step 2: Obtain an initial basic feasible solution using any one of the following methods.

(i) North – West Corner Method

(ii) Least Cost Method

(iii) Vogel's Approximation (or Penalty) Method

Step 3: Test the initial solution obtained in Step 2 for optimality using Modified Distribution (MODI) method (Also called u– v method). If the current solution is optimal, then stop. Otherwise determine a new improved solution.

Step 4: Repeat Step 3 until an optimal solution is reached.

5.4.1 North West Corner Method: Algorithm

Step 1: Start with the cell at the upper left (north – west) corner of the transportation matrix and allocate as much as possible equal to the minimum of the rim values for the first row and first column, i.e. $min\ \{a_1, b_1\}$.

Step 2: (a) If allocation made in Step 1 is equal to the supply available then move vertically down to the next row and same column and apply Step 1 again, for next allocation.

(b) If allocation made in Step 1 is equal to the demand of the destination then move horizontally to the next column and same row and apply Step 1 again, for next allocation.

(c) If allocation is equal to the demand and supply available then move diagonally, for the next allocation.

Step 3: Continue the procedure step by step till an allocation is made in the south – east corner cell of the transportation table.

Remark: If during the process of making allocation at a particular cell, supply equals demand, then next allocation of magnitude zero can be made in a cell either in the next row or column. This condition is known as degeneracy.

5.4.2 Problems on North – West Corner Method

Illustration 5.1:

A company has four production units (sources) S_1, S_2, S_3 and S_4 at different places. The product manufactured at these places is to be distributed at 3 different destinations D_1, D_2 and D_3. The cost (in Rs.) of transportation per unit between sources and destinations are given in the following table (cost matrix). It also shows the available quantity (supply) of item at each sources and the required quantity (demand) of item at each destinations. Apply North – West Corner method to find initial feasible solution of this transportation problem.

Table 5.2

Destination→ Source	D_1	D_2	D_3	Supply
S_1	2	4	5	20
S_2	3	4	3	40
S_3	2	3	6	30
S_4	5	2	9	10
Demand	30	50	20	

Solution:

From the given transportation table (Table 5.2), it is clear that Total Supply = Total Demand = 100. So the problem is balanced transportation problem.

We start allocation from the North – West corner (upper – left corner) cell i.e. (1, 1) cell or (S_1, D_1) cell in the transportation table and allocate here 20 units which is the minimum of available supply 20 for S_1 and available demand 30 for D_1.

Table 5.3

	D_1	D_2	D_3	Supply
S_1	2 / 20	4 / 0	5 / 0	20
S_2	3	4	3	40
S_3	2	3	6	30
S_4	5	2	9	10
Demand	30	50	20	

Since the supply for source S_1 is 20 units, we cannot supply more items from this source. Hence we allocate 0 to all other cells in the row of source S_1. For next allocation we move to the next cell of the first column i.e. (S_2, D_1) cell and allocate as much as possible. The maximum units which can be allocated in this cell is 10. This meets the complete demand of D_1. So we allocate 0 to all other cells of this column. Still supply of source S_2 is not satisfied. So, we move to the next cell of the second row i.e. (S_2, D_2) cell and allocate as much as possible. Here we can allocate 30 units which is the minimum of available units 50 in the demand of destination D_2 and the remaining available units 30 in the supply of source S_2.

Table 5.4

	D_1	D_2	D_3	Supply
S_1	2 / 20	4 / 0	5 / 0	20
S_2	3 / 10	4 / 30	3 / 0	40
S_3	2 / 0	3	6	30
S_4	5 / 0	2	9	10
Demand	30	50	20	

We allocate 0 to the remaining cells of source S_2 because we complete the supply of S_2. But still there are 20 units left at destination D_2. So, we move to the cell (S_3, D_2) and here we can see that the available demand of destination D_2 is less than the available source 30 of the source

S_3. Thus, we allocate 20 units in the (S_3, D_2) cell and allocate 0 to the remaining cells of this column. We allocate remaining 10 units of supply at source S_3 in the cell (S_3, D_3).

Table 5.5

	D_1	D_2	D_3	Supply
S_1	2 [20]	4 [0]	5 [0]	20
S_2	3 [10]	4 [30]	3 [0]	40
S_3	2 [0]	3 [20]	6 [10]	30
S_4	5 [0]	2 [0]	9 [10]	10
Demand	30	50	20	

Finally, the remaining available demand of the destination D_3 i.e. 10 units is same as the supply at source S_4. So we allocate this 10 units to the cell (S_4, D_3). The resulting feasible solution is shown in the Table 5.5.

The total transportation cost of this initial feasible solution by North – West Corner Method is calculated as:

Total Cost = 20×2 + 10×3 + 30×4 + 20×3 + 10×6 + 10×9 = 400 Rs.

Illustration 5.2:

Find initial basic feasible solution of the following transportation problem using North – West Corner Method.

Table 5.6

Destination→ / Source	D_1	D_2	D_3	D_4	Supply
S_1	19	20	50	10	7
S_2	70	30	40	60	9
S_3	40	8	70	20	18
Demand	5	8	7	14	

Solution:

From the given transportation table (Table 5.6), it is clear that Total Supply = Total Demand = 34. So the problem is balanced transportation problem.

Table 5.7

	D_1	D_2	D_3	D_4	Supply
S_1	19 5	20 2	50 0	10 0	7
S_2	70 0	30 6	40	60	9
S_3	40 0	8 0	70	20	18
Demand	5	8	7	14	

We start allocation from the North – West corner (upper – left corner) cell i.e. (S_1, D_1) cell in the transportation table and allocate there 5 units which is the minimum of available supply 7 for S_1 and available demand 5 for D_1. This allocation fulfills the demand at destination D_1.

Hence we allocate 0 in all other cells of this column. We allocate the remaining 2 units of supply in the next cell of the first row i.e. (S_1, D_2) cell. This completes the total supply 7 at the source S_1. So we allocate 0 in all other cells of this row. But still the requirement of the destination D_2 does not meet so we move to the cell (S_2, D_2) and allocate the remaining demand of 6 units in this cell and 0 in the remaining cell of this column.

The remaining available supply at source S_2 is 3 units, but the demand at destination D_3 is 7 units. So we allocate 3 units at the cell (S_2, D_3). This allocation leads us to allocate 0 in the cell (S_2, D_4). The remaining 4 units of demand for the destination D_3 are allocated at the next cell (S_3, D_3). Finally, the demand of 14 units at destination D_4 can be meet by allocating 14 units at the cell (S_3, D_4). The resulting feasible solution is shown in the Table 5.8. One should check the sum of allocation in rows and columns.

Table 5.8

	D_1	D_2	D_3	D_4	Supply
S_1	19 5	20 2	50 0	10 0	7
S_2	70 0	30 6	40 3	60 0	9
S_3	40 0	8 0	70 4	20 14	18
Demand	5	8	7	14	

The total transportation cost of this initial feasible solution by North – West Corner Method is calculated as:

Total Cost = 5×19 + 2×20 + 6×30 + 3×40 + 4×70 + 14×20 = 995 Rs.

5.4.3 Least Cost Method: Algorithm

Step 1: Select the cell with the lowest unit cost in the entire transportation table and allocate as much as possible to this cell and eliminate that row or column in which either supply or demand is exhausted. If both row and a column are satisfied simultaneously, only one may be crossed out.

In case the smallest unit cost cell is not unique, then select the cell where maximum allocation can be made.

Step 2: After adjusting the supply and demand for all uncrossed-out rows and columns repeat the procedure with the next lowest unit cost among the remaining rows and columns of the table and allocate as much as possible to this cell and eliminate that row and column in which either supply or demand is exhausted.

Step 3: Repeat the procedure until the entire available supply at various sources and demand at various destinations is satisfied.

5.4.4 Problems on Least Cost Method

Illustration 5.3:

Find initial feasible solution of the following transportation problem using Least Cost Method:

Table 5.9

Destination→ / Source	D_1	D_2	D_3	Supply
S_1	6	4	1	50
S_2	3	8	7	40
S_3	4	4	2	60
Demand	20	95	35	

Solution:

The lowest unit cost in the entire transportation table (Table 5.9) is 1 and it is found at the cell (S_1, D_3). So we make first allocation in this cell. We allocate $\min\{50,35\} = 35$ units in this cell as shown in Table 5.10.

Table 5.10

	D_1	D_2	D_3	Supply
S_1	6	4	1 35	50
S_2	3	8	7 0	40
S_3	4	4	2 0	60
Demand	20	95	35	

This will fulfill the demand at the destination D_3 and leave the supply of 15 units at the source S_1. Thus, the reduced transportation table can be formed as shown in Table 5.11.

Table 5.11

	D_1	D_2	Supply
S_1	6 0	4	15
S_2	3 20	8	40
S_3	4 0	4	60
Demand	20	95	

Repeating the process of allocation, we find that the next lowest cost is 3 at cell is (S_2, D_1) and possible allocation for this cell is $\min\{supply, demand\} = \min\{40, 20\} = 20$ units as shown in Table 5.11. This will fulfill the demand of the destination D_1 and leave the supply of 20 units at the source S_2. The reduced transportation table and allocation of remaining units is shown in Table 5.12.

Table 5.12

	D_2	Supply
S_1	4 15	15
S_2	8 20	20
S_3	4 60	60
Demand	95	

Since supply at each sources and demand at each destination is exhausted, the solution is arrived, and it is shown in Table 5.13.

Table 5.13

	D_1	D_2	D_3	Supply
S_1	6 [0]	4 [15]	1 [35]	50
S_2	3 [20]	8 [20]	7 [0]	40
S_3	4 [0]	4 [60]	2 [0]	60
Demand	20	95	35	

The total transportation cost of this initial feasible solution by Least Cost Method is calculated as:

Total Cost = 15×4 + 35×1 + 20×3 + 20×8 + 60×4 = 555 Rs.

Illustration 5.4:

Solve the transportation problem given in Illustration 5.2 using Least Cost Method and find initial basic feasible solution.

Solution:

The problem given in Illustration 5.2 is

Table 5.14

Destination→ Source	D_1	D_2	D_3	D_4	Supply
S_1	19	20	50	10	7
S_2	70	30	40	60	9
S_3	40	8	70	20	18
Demand	5	8	7	14	

The lowest unit cost in the entire transportation table is 8 at the cell (S_3, D_2). So allocate $\min\{supply, demand\} = \min\{18, 8\} = 8$ units at this cell. This allocation fulfill the demand of destination D_2 and leave 10 units at the source S_3. Hence allocate 0 in all other cells of the column for destination D_2. Next lowest cost among non allocated (unoccupied) cells is 10 at the cell (S_1, D_4) and possible allocation at this cell is 7 units. This allocation fulfill the supply at the source S_1 but leave the demand requirement of 7 units at the destination D_4. Hence allocate 0 in all other non allocated cells of the row for source S_1. The next lowest cost among non allocated cells of the transportation table is 20 at (S_3, D_4) cell. Allocate as much as units possible for this cell i.e. 7 units. This meets the complete demand at the destination D_4 and leave the availability of 3 units at the source S_3. So allocate 0 to all other non allocated cells of the column for destination D_4. Repeating this process, we find the next allocation sequence as: 7 units at (S_2, D_3) cell with 0 at (S_3, D_3) cell, 3 units at (S_3, D_1) cell and finally 2 units at (S_2, D_1) cell.

The solution is shown in Table 5.15.

Table 5.15

	D_1	D_2	D_3	D_4	Supply
S_1	19 0	20 0	50 0	10 7	7
S_2	70 2	30 0	40 7	60 0	9
S_3	40 3	8 8	70 0	20 7	18
Demand	5	8	7	14	

The total transportation cost of this initial solution by Least Cost Method is calculated as:

Total Cost = 7×10 + 2×70 + 7×40 + 3×40 + 8×8 + 7×20

= 70 + 140 + 280 + 120 + 64 + 140

= 814 Rs.

5.4.5 Vogel's Approximation Method: Algorithm

Step 1: Calculate penalties for each row (column) by taking the difference between the smallest and next smallest unit transportation cost in the same row (column). (Note: This difference indicates the penalty or extra cost which has to be pied if one fails to allocate to the cell with the minimum unit transportation cost.)

Step 2: Select the row or column with the largest penalty and allocate as much as possible in the cell having the least cost in the selected row or column satisfying the rim conditions. If there is a tie in the values of penalties, it can be broken by selecting the cell where maximum allocation can be made.

Step 3: Adjust the supply and demand and cross out the satisfied row or column. If a row and a column are satisfied simultaneously, only one of them is crossed out and the remaining row (column) is assigned a zero supply (demand). Any row or column with zero supply or demand should not be used in computing future penalties.

Step 4: Repeat Steps 1 to 3 until the entire available supply at various sources and demand at various destinations are satisfied.

5.4.6 Problems on Vogel's Approximation Method

Illustration 5.5:

Find initial feasible solution of the following transportation problem (Illustration 5.2) using Vogel's Approximation Method:

Table 5.16

Destination→ Source	D_1	D_2	D_3	D_4	Supply
S_1	19	20	50	10	7
S_2	70	30	40	60	9
S_3	40	8	70	20	18
Demand	5	8	7	14	

Solution:

We calculate differences (penalty costs) of the lowest and the second lowest cost for each row and column as shown in Table 5.17. Observe that the maximum penalty, 21 occurs in column D_1. Thus, we first allocate in the cell of column D_1 with minimum cost i.e. the cell (S_1, D_1). The maximum possible allocation in this cell is 5 units. This allocation fulfills demand of column D_1 and leaves the supply of 2 units at row S_1. Hence we allocate 0 in all other cells of column D_1.

Table 5.17

	D_1	D_2	D_3	D_4	Supply	Differences
S_1	19 [5]	20	50	10	7	9
S_2	70 [0]	30	40	60	9	10
S_3	40 [0]	8	70	20	18	12
Demand	5	8	7	14		
Differences	**21**	12	10	10		

The reduced transportation table with the calculation of new penalty costs for each row and column is shown in Table 5.18.

The maximum penalty, 12 occurs in column D_2 as well as row S_3. Thus, we allocate in the cell with minimum cost of column D_2 or row S_3 i.e. the cell (S_3, D_2). The maximum possible allocation in this cell is 8 units. This allocation fulfills demand of column D_2 and leaves the supply of 10 units at row S_3. Hence we allocate 0 in all other non allocated cells of column D_2.

Table 5.18

	D_2	D_3	D_4	Supply	Differences
S_1	20 / 0	50	10	2	10
S_2	30 / 0	40	60	9	10
S_3	8 / 8	70	20	18	**12**
Demand	8	7	14		
Differences	**12**	10	10		

The reduced transportation table with the calculation of new penalty costs for each row and column is shown in Table 5.19.

The maximum penalty, 50 occurs in row S_3. Thus, we allocate in the cell of row S_3 with minimum cost i.e. the cell (S_3, D_4). The maximum possible allocation in this cell is 10 units. This allocation fulfills supply of row S_3 and leaves the demand of 4 units at column D_4. Hence we allocate 0 in all other non allocated cells of row S_3.

Table 5.19

	D_3	D_4	Supply	Differences
S_1	50	10	2	40
S_2	40	60	9	20
S_3	70 / 0	20 / 10	10	**50**
Demand	7	14		
Differences	10	10		

The reduced transportation table with the calculation of penalty costs for each row and column is shown in Table 5.20. The remaining allocation is done as per the algorithm of Vogel's Approximation Method and it is shown in Table 5.20. The final resulting initial solution is shown in Table 5.21.

Table 5.20

	D_3	D_4	Supply	Differences
S_1	50 $\boxed{0}$	10 $\boxed{2}$	2	40
S_2	40 $\boxed{7}$	60 $\boxed{2}$	9	20
Demand	7	4		
Differences	10	**50**		

Table 5.21

	D_1	D_2	D_3	D_4	Supply
S_1	19 $\boxed{5}$	20 $\boxed{0}$	50 $\boxed{0}$	10 $\boxed{2}$	7
S_2	70 $\boxed{0}$	30 $\boxed{0}$	40 $\boxed{7}$	60 $\boxed{2}$	9
S_3	40 $\boxed{0}$	8 $\boxed{8}$	70 $\boxed{0}$	20 $\boxed{10}$	18
Demand	5	8	7	14	

The total transportation cost of this initial feasible solution by Vogel's Approximation Method is calculated as:

Total Cost = $5\times19 + 2\times10 + 7\times40 + 2\times60 + 8\times8 + 10\times20 = 779$ Rs.

Illustration 5.6:

Find solution of the following transportation problem using Vogel's Approximation Method:

Table 5.22

Destination→ / Source	D_1	D_2	D_3	D_4	D_5	D_6	Supply
S_1	2	1	5	5	2	7	40
S_2	5	2	2	6	5	6	30
S_3	5	7	6	2	6	1	55
S_4	6	2	2	1	2	2	25
Demand	25	45	15	35	25	5	

Solution:

We calculate differences (penalty costs) of the smallest and the second smallest cost for each row and column as shown in Table 5.23. Observe that the maximum penalty, 3 occurs in three columns D_1, D_3, D_5 and one row S_2. If we select column D_1 for allocation, the cell with minimum cost in the column D_1 is the cell (S_1, D_1) and we can allocate 25 units there. If we select column D_3 or D_5 or row S_2, it is observed that we can allocate less than or equals 25 units. So we choose the column D_1 and the cell (S_1, D_1). This allocation satisfies demand of column D_1 and leaves the supply of 15 units at row S_1. Hence we allocate 0 in all other cells of column D_1.

Table 5.23

	D_1	D_2	D_3	D_4	D_5	D_6	Supply	Differences
S_1	2 / 25	1	5	5	2	7	40	1
S_2	5 / 0	2	2	6	5	6	30	3
S_3	5 / 0	7	6	2	6	1	55	1
S_4	6 / 0	2	2	1	2	2	25	1
Demand	25	45	15	35	25	5		
Diff.	**3**	1	3	1	3	1		

Table 5.24

	D_2	D_3	D_4	D_5	D_6	Supply	Differences
S_1	1	5	5	2	7	15	1
S_2	2 / 30	2 / 0	6 / 0	5 / 0	6 / 0	30	**3**
S_3	7	6	2	6	1	55	1
S_4	2	2	1	2	2	25	1
Demand	45	15	35	25	5		
Diff.	1	3	1	3	1		

The reduced transportation table with the calculation of new penalty costs for each row and column is shown in Table 5.24.

Again the maximum penalty, 3 occurs in columns D_3, D_5 and row S_2. But by examining the possible allocation, it is found that the maximum allocation of 30 units is possible if we select the row S_2 and the cell (S_2, D_2). This allocation fulfills supply of row S_2 and leaves the demand of 15 units at column D_2. Hence we allocate 0 in all other non allocated cells of row S_2.

The reduced transportation table with the calculation of new penalty costs for each row and column is shown in Table 5.25. The maximum penalty, 4 occurs in column D_5. Minimum cost in this column is found at two cells (S_1, D_5) and (S_4, D_5). But maximum allocation of 25 units is possible at the cell (S_4, D_5). This allocation fulfills supply of row S_4 as well as demand of column D_5. Hence we allocate 0 in all other non allocated cells of row S_4 and column D_5.

Table 5.25

	D_2	D_3	D_4	D_5	D_6	Supply	Differences
S_1	1	5	5	2 / 0	7	15	1
S_3	7	6	2	6 / 0	1	55	1
S_4	2 / 0	2 / 0	1 / 0	2 / 25	2 / 0	25	1
Demand	15	15	35	25	5		
Diff.	1	3	1	**4**	1		

Table 5.26

	D_2	D_3	D_4	D_6	Supply	Differences
S_1	1 / 15	5 / 0	5 / 0	7 / 0	15	4
S_3	7 / 0	6	2	1	55	1
Demand	15	15	35	5		
Diff.	**6**	1	3	6		

The reduced transportation table with the calculation of new penalty costs for each row and column is shown in Table 5.26.

The maximum penalty, 6 occurs in columns D_2 and D_6. If we select column D_6, we can allocate 5 units only. So we choose column D_2 and the cell with minimum cost in this column is (S_1, D_2). Here we allocate as possible i.e. 15 units. This allocation fulfills both supply of row S_1 as well as demand of column D_2. Hence we allocate 0 in all other non allocated cells of row S_4 and column D_5.

Now there is only one row left with non allocated cells. So we can easily allocate. The final solution table is shown in Table 5.27.

Table 5.27

	D_1	D_2	D_3	D_4	D_5	D_6	Supply	Differences			
S_1	2 25	1 15	5 0	5 0	2 0	7 0	40	1	1	1	4
S_2	5 0	2 30	2 0	6 0	5 0	6 0	30	3	3	-	-
S_3	5 0	7 0	6 15	2 35	6 0	1 5	55	1	1	1	1
S_4	6 0	2 0	2 0	1 0	2 25	2 0	25	1	1	1	-
Demand	25	45	15	35	25	5					
Diff.	3	1	3	1	3	1					
	--	1	3	1	3	1					
	--	1	3	1	4	1					
	--	6	1	3	--	6					

The total transportation cost of this initial feasible solution by Vogel's Approximation Method is calculated as:

Total Cost = 25×2 + 15×1 + 30×2 + 15×6 + 35×2 + 5×1 + 25×2 = 340 Rs.

5.5 FINAL (OPTIMAL) SOLUTION: MODIFIED DISTRIBUTION (MODI) METHOD

5.5.1 Algorithm: Steps for Optimal Solution of Transportation Problem using MODI Method

The initial solution can be obtained by any of the three methods discussed earlier (NWCM, LCM or VAM).

Step 1: For an initial basic feasible solution with $m + n - 1$ occupied cells, calculate u_i and v_j for rows and columns using the relationship

$$u_i + v_j = c_{ij} \qquad \text{for each occupied cell } (i, j).$$

Note: To start with, any one of $u_i's$ or $v_j's$ is assigned the value zero. It is better to assign zero for a particular u_i or v_j where there are maximum number of allocations in a row or column respectively, as it will reduce arithmetic work considerable.

Step 2: For unoccupied cells, calculate opportunity cost by using the relationship

$$d_{ij} = c_{ij} - (u_i + v_j)$$

Step 3: Examine sign of each d_{ij}

(i) If all $d_{ij} > 0$, then current basic feasible solution is optimal.

(ii) If $d_{ij} = 0$, then current basic feasible solution will remain unaffected but an alternative solution exists.

(iii) If one or more $d_{ij} < 0$, then an improved solution can be obtained by entering unoccupied cell (i, j) in the basis. An unoccupied cell having the largest negative value of d_{ij} is chosen for entering into the solution mix.

Step 4: Construct a closed path (or loop) for the unoccupied cell with largest negative opportunity cost. Start the closed path with the selected unoccupied cell and mark a plus sign (+) in this cell, trace a path along the rows (or columns) to an occupied cell, mark the corner with minus sign (–) and continue down the column (or row) to an occupied cell and mark the corner with plus sign (+) and minus sign (–) alternatively. Close the path back to the selected unoccupied cell.

Step 5: Select the smallest quantity amongst the cells marked with minus sign on the corners of closed loop. Allocate this value to the selected unoccupied cell and add it to other occupied cells marked with plus signs and subtract it from the occupied cells marked with minus signs.

Step 6: Obtain a new improved solution by allocating units to the unoccupied cell according to Step 5 and calculate the new total transportation cost.

Step 7: Test the revised solution further for optimality. The procedure terminates when all $d_{ij} \geq 0$, for unoccupied cells.

5.5.2 Problems on MODI Method

Illustration 5.7:

Find optimal solution of transportation problem by applying MODI method using the data of Illustration 5.2 i.e.

Table 5.28

Destination→ / Source	D_1	D_2	D_3	D_4	Supply
S_1	19	20	50	10	7
S_2	70	30	40	60	9
S_3	40	8	70	20	18
Demand	5	8	7	14	

Solution:

First obtain initial basic feasible solution of the given transportation problem using any one of the three methods viz. North – West Corner Method, Least Cost Method and Vogel's Approximation Method.

Applying Vogel's Approximation Method as mentioned in Illustration 5.2, we obtain allocation in an initial basic feasible solution of the given problem with the total transportation cost Rs. 779 is as follow:

Table 5.29

	D_1	D_2	D_3	D_4	Supply
S_1	19 (5)	20	50	10 (2)	7
S_2	70	30	40 (7)	60 (2)	9
S_3	40	8 (8)	70	20 (10)	18
Demand	5	8	7	14	

By examining Table 5.29, number of row m = 3, number of column n = 4 and number of occupied cells = m + n – 1 = 6. Hence the modified distribution method can be applied to obtain optimum solution. Let us first do the optimality test for this solution.

Assign the value zero to any one of u_i $(i = 1,2 \ldots, m)$ or v_j $(j = 1,2, \ldots, n)$. Since maximum number of occupied cells i.e. 3 occupied cells found in fourth column, we can initially assign the value zero to v_4 as shown in Table 5.30. Now apply the formula $C_{ij} = u_i + v_j$ for all occupied cells, where C_{ij} denotes the cost in the cell (i, j), we get

$$C_{14} = u_1 + v_4 \quad or \quad 10 = u_1 + v_4 \quad (1)$$

$$C_{24} = u_2 + v_4 \quad or \quad 60 = u_2 + v_4 \quad (2)$$

$$C_{34} = u_3 + v_4 \quad or \quad 20 = u_3 + v_4 \quad (3)$$

$$C_{11} = u_1 + v_1 \quad or \quad 19 = u_1 + v_1 \quad (4)$$

$$C_{23} = u_2 + v_3 \quad or \quad 40 = u_2 + v_3 \quad (5)$$

$$C_{32} = u_3 + v_2 \quad or \quad 8 = u_3 + v_2 \quad (6)$$

Substituting $v_4 = 0$ in Equations (1), (2) and (3), we get $u_1 = 10$, $u_2 = 60$ and $u_3 = 20$. Replacing these values of u_1, u_2 and u_3 respectively in Equations (4), (5) and (6), we obtain $v_1 = 9, v_3 = -20$ and $v_2 = -12$.

Table 5.30

	D_1	D_2	D_3	D_4	Supply	u_i
S_1	19 [5]	20 [ ]	50 [ ]	10 [2]	7	$u_1 = 10$
S_2	70 [ ]	30 [ ]	40 [7]	60 [2]	9	$u_2 = 60$
S_3	40 [ ]	8 [8]	70 [ ]	20 [10]	18	$u_3 = 20$
Demand	5	8	7	14		
v_j	$v_1 = 9$	$v_2 = -12$	$v_3 = -20$	$v_4 = 0$		

Now, calculate opportunity cost $d_{ij} = C_{ij} - (u_i + v_j)$ for all unoccupied cells.

$$d_{12} = C_{12} - (u_1 + v_2) = 20 - (10 - 12) = +22$$

$$d_{13} = C_{13} - (u_1 + v_3) = 50 - (10 - 20) = +60$$

$$d_{21} = C_{21} - (u_2 + v_1) = 70 - (60 + \quad 9) = +1$$

$$d_{22} = C_{22} - (u_2 + v_2) = 30 - (60 - 12) = -18$$

$$d_{31} = C_{31} - (u_3 + v_1) = 40 - (20 + \ 9) = +11$$

$$d_{33} = C_{33} - (u_3 + v_3) = 70 - (20 - 20) = +70$$

Since not all $d_{ij} \geq 0$, according to the optimality test condition the solution in Table 5.30 is not optimal solution. The value $d_{22} = -18$ indicates that the total transportation cost can be reduced in the multiple of 18 by shifting an allocation to this cell.

Choose the cell with negative opportunity cost i.e. the cell (S_2, D_2). Trace a closed – loop along the column D_2 to an occupied cell (S_3, D_2). Place a plus (+) sign in the cell (S_2, D_2) and minus (–) sign in the cell (S_3, D_2). Now take a right angle turn and locate an occupied cell (S_3, D_4). Place a plus (+) sign in this cell. Again take right angle turn up and locate an occupied cell (S_2, D_4). Place a minus (–) sign in this cell. At last, take a right angle turn on left side and complete the closed – loop as shown in Table 5.31.

Table 5.31

	D_1	D_2	D_3	D_4	Supply	u_i
S_1	19 5	20 +22	50 +60	10 2	7	$u_1 = 10$
S_2	70 +1	30 (+) –18	40 7	60 (–) 2	9	$u_2 = 60$
S_3	40 +11	8 (–) 8	70 +70	20 (+) 10	18	$u_3 = 20$
Demand	5	8	7	14		
v_j	$v_1 = 9$	$v_2 = -12$	$v_3 = -20$	$v_4 = 0$		

Now, examine the corner cells of the closed – loop with minus sign and select the smallest allocation value among them. From Table 5.31, we see that the allocation of 2 units at the cell (S_2, D_4) is the minimum allocation. Now, add this quantity (i.e. 2) to all the cells on the corner points of the closed path marked with plus signs and subtract it from those cells marked with minus signs. In this way, an unoccupied cell becomes an occupied cell. The modified solution is shown in Table 5.32. Again test the optimality of this revised solution in the same way as discussed earlier. The values of $u's$, $v_j's$ and $d_{ij}'s$ are shown in Table 5.32.

Table 5.32

	D_1	D_2	D_3	D_4	Supply	u_i
S_1	19 5	20 +22	50 +42	10 2	7	$u_1 = 10$
S_2	70 +19	30 2	40 7	60 +18	9	$u_2 = 42$
S_3	40 +11	8 6	70 +52	20 12	18	$u_3 = 20$
Demand	5	8	7	14		
v_j	$v_1 = 9$	$v_2 = -12$	$v_3 = -2$	$v_4 = 0$		

Since all $d_{ij} \geq 0$, the solution in Table 5.32 is optimal solution and the total transportation cost is calculated as:

Total Cost = 5×19 + 2×10 + 2×30 + 7×40 + 6×8 + 12×20

= 95 + 20 + 60 + 280 + 48 + 240 = 743 Rs.

Illustration 5.8:

Use MODI method to find optimal solution of the transportation problem given in Illustration 5.3 i.e.

Table 5.33

Destination→ Source	D_1	D_2	D_3	Supply
S_1	6	4	1	50
S_2	3	8	7	40
S_3	4	4	2	60
Demand	20	95	35	

Solution:

First obtain initial basic feasible solution of the given transportation problem using any one of the three methods viz. North – West Corner Method, Least Cost Method and Vogel's Approximation Method.

As illustrated in Illustration 5.3, the initial basic feasible solution using Least Cost Method is obtained as

Table 5.34

	D_1	D_2	D_3	Supply
S_1	6	4 [15]	1 [35]	50
S_2	3 [20]	8 [20]	7	40
S_3	4	4 [60]	2	60
Demand	20	95	35	

Total Cost = 15×4 + 35×1 + 20×3 + 20×8 + 60×4 = Rs. 555.

In Table 5.34, number of row m = 3, number of column n = 4 and number of occupied cells = m + n – 1 = 6. Hence the modified distribution method can be applied to obtain optimum solution.

Assign the value zero to any one of u_i $(i = 1,2,3)$ or v_j $(j = 1,2,3,4)$. Since maximum number of occupied cells i.e. 3 occupied cells found in second column, we can set $v_2 = 0$ as shown in Table 5.35. Now apply the formula $C_{ij} = u_i + v_j$ for all occupied cells, we get

$$C_{12} = u_1 + v_2 \quad or \quad 4 = u_1 + v_2 \quad (1)$$

$$C_{13} = u_1 + v_3 \quad or \quad 1 = u_1 + v_3 \quad (2)$$

$$C_{21} = u_2 + v_1 \quad or \quad 3 = u_2 + v_1 \quad (3)$$

$$C_{22} = u_2 + v_2 \quad or \quad 8 = u_2 + v_2 \quad (4)$$

$$C_{32} = u_3 + v_2 \quad or \quad 4 = u_3 + v_2 \quad (5)$$

Table 5.35

	D_1	D_2	D_3	Supply	u_i
S_1	6	4 [15]	1 [35]	50	$u_1 = 4$
S_2	3 [20]	8 [20]	7	40	$u_2 = 8$
S_3	4	4 [60]	2	60	$u_3 = 4$
Demand	20	95	35		
v_j	$v_1 = -5$	$v_2 = 0$	$v_3 = -3$		

Substituting $v_2 = 0$ in Equations (1), (4) and (5), we get $u_1 = 4, u_2 = 8$ and $u_3 = 4$. Replacing $u_2 = 8$ in Equation (3), we obtain $v_1 = -5$. Similarly by replacing $u_1 = 4$ in Equation (2), we get $v_3 = -3$.

Now, calculate opportunity cost $d_{ij} = C_{ij} - (u_i + v_j)$ for all unoccupied cells.

$$d_{11} = C_{11} - (u_1 + v_1) = 6 - (4 - 5) = +7$$

$$d_{23} = C_{23} - (u_2 + v_3) = 7 - (8 - 3) = +2$$

$$d_{31} = C_{31} - (u_3 + v_1) = 4 - (4 - 5) = +5$$

$$d_{33} = C_{33} - (u_3 + v_3) = 2 - (4 - 3) = +1$$

Since all $d_{ij} \geq 0$, the solution in Table 5.35 is optimal solution and the total transportation cost is Rs. 555.

Illustration 5.9:

Following table shows allocation of units in a transportation problem using Least Cost Method. Find the optimum solution of this transportation problem using MODI Method:

Table 5.36

	D_1	D_2	D_3	D_4	Supply
S_1	15	25	50	10 [10]	10
S_2	65 [3]	30	35 [9]	60	12
S_3	40 [4]	5 [10]	70	20 [4]	18
Demand	7	10	9	14	

Solution:

The total transportation cost of the given initial solution by Least Cost Method is calculated as:

Total Cost = 10×10 + 3×65 + 9×35 + 4×40 + 10×5 + 4×20 = 900 Rs.

We first set $u_3 = 0$ and applying the formula $C_{ij} = u_i + v_j$ for all occupied cells, we get

$$C_{14} = u_1 + v_4 \qquad or \qquad 10 = u_1 + v_4 \qquad (1)$$

$$C_{21} = u_2 + v_1 \qquad or \qquad 65 = u_2 + v_1 \qquad (2)$$

$$C_{23} = u_2 + v_3 \qquad or \qquad 35 = u_2 + v_3 \qquad (3)$$

$$C_{31} = u_3 + v_1 \qquad or \qquad 40 = u_3 + v_1 \qquad (4)$$

$$C_{32} = u_3 + v_2 \qquad or \qquad 5 = u_3 + v_2 \qquad (5)$$

$$C_{34} = u_3 + v_4 \qquad or \qquad 20 = u_3 + v_4 \qquad (6)$$

Substitute $u_3 = 0$ in Equations (4), (5) and (6) to get $v_1 = 40, v_2 = 5$ and $v_4 = 20$. Replacing $v_1 = 40$ in Equation (2), we obtain $u_2 = 25$. Similarly by replacing $v_4 = 20$ in Equation (1) and $u_2 = 25$ in Equation (3), we get $u_1 = -10$ and $v_3 = 10$ as shown in Table 5.37.

Table 5.37

	D_1	D_2	D_3	D_4	Supply	u_i
S_1	15	25	50	10 [10]	10	$u_1 = -10$
S_2	65 [3]	30	35 [9]	60	12	$u_2 = 25$
S_3	40 [4]	5 [10]	70	20 [4]	18	$u_3 = 0$
Demand	7	10	9	14		
v_j	$v_1 = 40$	$v_2 = 5$	$v_3 = 10$	$v_4 = 20$		

Now, calculate opportunity cost $d_{ij} = C_{ij} - (u_i + v_j)$ for all unoccupied cells.

$$d_{11} = C_{11} - (u_1 + v_1) = 15 - (-10 + 40) = -30$$

$$d_{12} = C_{12} - (u_1 + v_2) = 25 - (-10 + 5) = +30$$

$$d_{13} = C_{13} - (u_1 + v_3) = 50 - (-10 + 10) = +50$$

$$d_{22} = C_{22} - (u_2 + v_2) = 30 - (25 + 5) = 0$$

$$d_{24} = C_{24} - (u_2 + v_4) = 60 - (25 + 20) = +15$$

$$d_{33} = C_{33} - (u_3 + v_3) = 70 - (0 + 10) = +60$$

Since not all $d_{ij} \geq 0$, the solution in Table 5.37 is not optimal solution according to the optimality test condition. The value $d_{11} = -30$ indicates that the total transportation cost can be reduced in the multiple of 30 by shifting an allocation to this cell.

Choose the cell with negative opportunity cost i.e. the cell (S_1, D_1). Trace a closed – loop along the row S_1 to an occupied cell (S_3, D_1). Place a plus (+) sign in the cell (S_1, D_1) and minus (–)

sign in the cell (S_3, D_1). Now take a right angle turn and locate an occupied cell (S_3, D_4). Place a plus (+) sign in this cell. Again take right angle turn up and locate an occupied cell (S_1, D_4). Place a minus (–) sign in this cell. At last, take a right angle turn on left side and complete the closed – loop as shown in Table 5.38.

Table 5.38

	D_1	D_2	D_3	D_4	Supply	u_i
S_1	15 (+) –30	25 +30	50 +50	10 (–) 10	10	$u_1 = -10$
S_2	65 3	30 +0	35 9	60 +15	12	$u_2 = 25$
S_3	40 (–) 4	5 10	70 +60	20 (+) 4	18	$u_3 = 0$
Demand	7	10	9	14		
v_j	$v_1 = 40$	$v_2 = 5$	$v_3 = 10$	$v_4 = 20$		

Now, examine the corner cells of the closed – loop with minus sign and select the smallest allocation value among them. From Table 5.38, we see that the allocation of 4 units at the cell (S_3, D_1) is the minimum allocation. Now, add this quantity (i.e. 4) to all the cells on the corner points of the closed path marked with plus signs and subtract it from those cells marked with minus signs. In this way, an unoccupied cell becomes an occupied cell. The modified solution is shown in Table 5.39. Again test the optimality of this revised solution in the same way as discussed earlier. The values of $u_i's$, $v_j's$ and $d_{ij}'s$ are shown in Table 5.39.

Table 5.39

	D_1	D_2	D_3	D_4	Supply	u_i
S_1	15 4	25 +30	50 +65	10 6	10	$u_1 = 0$
S_2	65 3	30 –15	35 9	60 +0	12	$u_2 = 50$
S_3	40 +15	5 10	70 +75	20 8	18	$u_3 = 10$
Demand	7	10	9	14		
v_j	$v_1 = 15$	$v_2 = -5$	$v_3 = -15$	$v_4 = 10$		

The total transportation cost of this improved solution is calculated as:

Total Cost = 15×4 + 10×6 + 65×3 + 35×9 + 5×10 + 20×8 = 840 Rs.

Since not all $d_{ij} \geq 0$, the solution in Table 5.39 is not optimal solution. The value $d_{22} = -15$ indicates that the total transportation cost can be reduced in the multiple of 15 by shifting an allocation to this cell.

Choose the cell (S_2, D_2) with negative opportunity cost. Trace a closed – loop along the row S_2 to an occupied cell (S_3, D_2). Place a plus (+) sign in the cell (S_2, D_2) and minus (–) sign in the cell (S_3, D_2). Now take a right angle turn and locate an occupied cell (S_3, D_4). Place a plus (+) sign in this cell. Again take right angle turn up and locate an occupied cell (S_1, D_4). Place a minus (–) sign in this cell. Now take a right angle turn on left side to locate an occupied cell (S_1, D_1) and put a plus (+) sign in the cell. Move down towards an occupied cell (S_2, D_1) and place minus (–) sign in this cell. Finally to complete the closed – loop move right upto the starting cell (S_2, D_2) as shown in Table 5.40.

Table 5.40

	D_1	D_2	D_3	D_4	Supply	u_i
S_1	15 (+) 4	25 +30	50 +65	10 (–) 6	10	$u_1 = 0$
S_2	65 (–) 3	30 (+) –15	35 9	60 +0	12	$u_2 = 50$
S_3	40 +15	5 (–) 10	70 +75	20 (+) 8	18	$u_3 = 10$
Demand	7	10	9	14		
v_j	$v_1 = 15$	$v_2 = -5$	$v_3 = -15$	$v_4 = 10$		

Now, examine the corner cells of the closed – loop with minus sign and select the smallest allocation value among them. From Table 5.40, we see that the allocation of 3 units at the cell (S_2, D_1) is the minimum allocation. Now, add this quantity (i.e. 3) to all the cells on the corner points of the closed path marked with plus signs and subtract it from those cells marked with minus signs. In this way, an unoccupied cell becomes an occupied cell. The modified solution is shown in Table 5.41. Again test the optimality of this revised solution in the same way as discussed earlier. The values of $u's$, $v_j's$ and $d_{ij}'s$ are shown in Table 5.41.

Table 5.41

	D_1	D_2	D_3	D_4	Supply	u_i
S_1	15 7	25 +30	50 +50	10 3	10	$u_1 = 0$
S_2	65 +15	30 3	35 9	60 +15	12	$u_2 = 35$
S_3	40 +15	5 7	70 +60	20 11	18	$u_3 = 10$
Demand	7	10	9	14		
v_j	$v_1 = 15$	$v_2 = -5$	$v_3 = 0$	$v_4 = 10$		

Since all $d_{ij} \geq 0$, the solution in Table 5.41 is optimal solution and the total transportation cost is calculated as:

Total Cost = 15×7 + 10×3 + 30×3 + 35×9 + 5×7 + 20×11

= 105 + 30 + 90 + 315 + 35 + 220 = 795 Rs.

5.6 UNBALANCED TRANSPORTATION PROBLEM

Illustration 5.10:

Find initial basic feasible solution of the following transportation problem using (i) North-West Corner Method and (ii) Least Cost Method:

Table 5.42

Destination→ Source	D_1	D_2	D_3	Supply
S_1	30	20	10	700
S_2	5	15	25	400
Demand	300	400	250	

Solution:

Here the Total Supply = 1100 unit and Total Demand = 950 unit. So, Total Supply G Total Demand. Hence the problem is unbalanced problem and we have to consider a dummy destination requiring 1100 – 950 = 150 units of quantity. Thus, the new transportation problem is

Table 5.43

Destination→ / Source	D_1	D_2	D_3	D_4	Supply
S_1	30	20	10	0	700
S_2	5	15	25	0	400
Demand	300	400	250	150	

(i) Solving this by North-West Corner Method, we get a initial basic feasible solution as shown in Table 5.44.

Table 5.44

	D_1	D_2	D_3	D_4	Supply
S_1	30 [300]	20 [400]	10	0	700
S_2	5	15	25 [250]	0 [150]	400
Demand	300	400	250	150	

The total transportation cost is calculated as:

Total Cost = 300×30 + 400×20 + 250×25 + 150×0 = 23250 Rs.

(ii) Using Least Cost Method, we get an initial basic feasible solution as in Table 5.45.

Table 5.45

	D_1	D_2	D_3	D_4	Supply
S_1	30	20 [300]	10 [250]	0 [150]	700
S_2	5 [300]	15 [100]	25	0	400
Demand	300	400	250	150	

The total transportation cost is calculated as:

Total Cost = 300×20 + 250×10 + 150×0 + 300×5 + 100×15 = 11500 Rs.

The alternative solution by Least Cost Method can be obtained as shown in Table 5.46.

Table 5.46

	D_1	D_2	D_3	D_4	Supply
S_1	30 / 50	20 / 400	10 / 250	0	700
S_2	5 / 250	15	25	0 / 150	400
Demand	300	400	250	150	

The total transportation cost is calculated as:

Total Cost = 50×30 + 400×20 + 250×10 + 250×5 + 150×0 = 13250 Rs.

❖ **CHECK YOUR PROGRESS**

- **Answer the following multiple choice questions.**

Que. 1 In which of the following method allocation starts from the cell with lowest cost?

(a) North–West corner (b) least cost method

(c) Vogel's approximation (d) none of these

Que. 2 Which of the following method is a method of penalty?

(a) North–West corner (b) least cost method

(c) Vogel's approximation (d) none of these

Que. 3 In a transportation problem, if total number of allocations = $m + n - 1$ where m is number of rows and n is number of columns, it means

(a) there is no degeneracy (b) problem is unbalanced

(c) problem is degenerate (d) solution is optimal

Que. 4 Which of the following method considers difference between two least cost for each row and column while finding initial basic feasible solution of a transportation problem?

(a) North–West corner (b) least cost

(c) row minima (d) Vogel's approximation

Que. 5 The dummy source or destination in a transportation problem is added to

(a) satisfy rim condition
(b) prevent solution from becoming degenerate

(c) ensure that total cost does not exceed a limit

(d) all of the above

Que. 6 Once the initial basic feasible solution is computed, what is the next step in transportation problem?

(a) applying VAM (b) applying optimality test

(c) constructing closed loop (d) none of the above

Que. 7 The large negative opportunity cost value in an unused cell in a transportation table ischosen to improve the current solution because

(a) it represents per unit cost reduction

(b) it represents per unit cost improvement

(c) it ensures no rim requirement violation

(d) none of the above

Que. 8 An alternative optimal solution to a minimization transportation problem exists whenever opportunity cost corresponding to unused route of transportation is:

(a) positive and greater than zero

(b) positive with at least one equal to zero

(c) negative with at least one equal to zero

(d) none of the above

Answer the following questions in brief.

Que. 1 What is transportation problem?

Que. 2 Write steps involved in North–West corner method.

Que. 3 State the mathematical formulation for transportation problem.

Que. 4 Write the steps for transportation algorithm.

Que. 5 State different methods to obtain the initial basic feasible solution of a transportation problem.

Que. 6 What is rim condition?

Que. 7 What do you mean by opportunity cost?

❖ Answer the following questions in detail.

Que. 1 Write steps for MODI method.

Que. 2 Three fertilizers factories X, Y and Z located at different places of the country produce 6, 4 and 5 lakh tones of urea respectively. Under the directive of the central government, they are to be distributed to 3 States A, B and C as 5, 3 and 7 lakh respectively. The transportation cost per tones in rupees is given below:

State → / Factory	A	B	C
X	11	17	16
Y	15	12	14
Z	20	12	15

Find out suitable transportation pattern at minimum cost by North West corner method.

Que. 3 Find out the initial basic feasible solution of the transportation problem given in Que. 2 by using least cost method.

Que. 4 Determine an initial basic feasible solution of the following transportation problem by using Vogel's approximation method.

Destination → / Source	D_1	D_2	D_3	D_4	Supply
S_1	6	4	1	5	14
S_2	8	9	2	7	16
S_3	4	3	6	2	5
Demand	6	10	15	4	

Que. 5 Obtain the optimal solution to the transportation problem given in Que. 4.

Que. 6 Obtain the optimal solution to the following transportation problem.

Destination→ / Source	D_1	D_2	D_3	D_4	Supply
S_1	2	3	11	7	6
S_2	1	0	6	1	1
S_3	5	8	15	10	10
Demand	7	5	3	2	

MBA

SEMESTER-1

QUANTITATIVE MANAGEMENT

BLOCK: 2

Authors' Name:	Dr. Umesh Raval Dr. Paresh Andhariya Dr. Hiren Patel
Review (Subject):	Prof. (Dr.) Manoj Shah Dr. Ravi Vaidya Dr. Maulik Desai
Review (Language):	Dr. Jainee Shah
Editor's Name:	Prof. (Dr.) Manoj Shah, Professor and Director, School of Commerce and Management, Dr. Babasaheb Ambedkar Open University, Ahmedabad.
Co-Editor's Name:	Dr. Dhaval Pandya Assistant Professor, School of Commerce and Management, Dr. Babasaheb Ambedkar Open University, Ahmedabad.

Publisher's Name:	Registrar, Dr. Babasaheb Ambedkar Open University, 'Jyotirrmay Parisar', opp. Shri Balaji Temple, Chharodi, Ahmedabad, 382481, Gujarat, India.
Edition:	2022 (First Edition)
ISBN:	978-93-5598-319-0

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means without permission in writing from Dr. Babasaheb Ambedkar Open University, Ahmedabad.

UNIT – 6	ASSIGNMENT MODEL

6.1 INTRODUCTION

Assignment problem arises in diverse situations, where one needs to determine an optimal way to assign subjects to subjects in the best possible way. Given n resources (or facilities) and n activities (or jobs), and effectiveness (in terms of cost, profit, time, etc.), of each resource (facility) for each activity (job), the problem lies in assigning each resource to one and only one activity (job) so that the given measure of effectiveness is optimized. An assignment problem is a special type of linear programming problem where the objective is to minimize the cost or time of completing a number of jobs by a number of persons.

6.2 IMPORTANCE

Problems related to assignment arise in a range of fields, for example, healthcare, transportation, education, and sports. For example, in assigning machines to factory orders, in assigning sales/marketing people to sales territories, in assigning contracts to bidders by systematic bid-evaluation, in assigning teachers to classes, in assigning accountants to accounts of the clients.

6.3 MATHEMATICAL MODEL OF AN ASSIGNMENT PROBLEM

The general data matrix for assignment problem is shown in the following table:

Table 6.1: Data Matrix

Resources (Workers)	Activities (Jobs)				Supply
	J_1	J_2	… …	J_4	
W_1	c_{11} x_{11}	c_{12} x_{12}	… … …	c_{1n} x_{1n}	1
W_2	c_{21} x_{21}	c_{22} x_{22}	… … …	c_{2n} x_{2n}	1
:	:	:	: :	:	:
W_n	c_{n1} x_{n1}	c_{n2} x_{n2}	… … …	c_{nn} x_{nn}	1
Demand	1	1	… …	1	N

x_{ij} denote the assignment of facility i to job j such that

$x_{ij} = 1$ if facility i is assigned to job j

$= 0$ otherwise.

The mathematical model of the assignment problem can be stated as:

Minimize $Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$

Subject to the constraints

$$\sum_{j=1}^{n} x_{ij} = 1, \quad \text{for all i} \qquad \sum_{i=1}^{n} x_{ij} = 1, \quad \text{for all j}$$

Thus, the mathematical model of assignment problem is a particular case of the transportation problem because (i) the cost matrix is a square matrix, and (ii) the optimal solution table for the problem would have only one assignment in a given row or a column.

In this way, the assignment problem is a special case of transportation problem.

6.4 BASIC ASSIGNMENT STEPS

Solution methods of Assignment Problem:

Various methods to solve an assignment problem are

(i)	Enumeration method	(ii)	Simplex method
(iii)	Transportation method	(iv)	Hungarian method

We will study the Hungarian method.

6.4.1 Hungarian Method

Algorithm:

Step 1: Develop the cost table from the given problem.

If the number of rows is not equal to the number of columns and vice versa, a dummy row or dummy column must be added. The assignment cost for dummy cells are always zero.

Step 2: Find the opportunity cost table

(a) Locate the smallest element in each row of the given cost table and then subtract that from each element of that row, and;

(b) In the reduced matrix obtained, locate the smallest element in each column and then subtract that from each element of that column.

Step 3: Make assignments in the opportunity cost matrix

(a) Examine rows successively until a row with exactly one unmarked zero is obtained. Make an assignment to this single zero by making a square around it.

(b) For each zero value that becomes assigned, eliminate (strike off) all other zeros in the same row and/or column.

(c) Repeat 3(a) and 3(b) for each column also with exactly single zero value cell that has not been assigned.

(d) If a row and/or column has two or more unmarked zeros and one cannot be chosen by inspection, then choose the assigned zero cell arbitrarily.

(e) Continue this process until all zeros in rows/columns are either enclosed (assigned) or struck off.

Step 4: Optimality criterion

If the number of assigned cells is equal to the number of rows/columns, then it is an optimal solution. The total cost associated with this solution is obtained by adding original cost figures in the occupied cells.

If a zero cell was chosen arbitrarily in Step 3, there exists an alternative optimal solution. But if no optimal solution is found, then go to Step 5.

Step 5: Revise the opportunity cost table

Draw a set of horizontal and vertical lines to cover all the zeros in the revised cost table obtained from Step 3, by using the following procedure:

(a) For each row in which no assignment was made, mark a tick.
(b) Examine the marked rows. If any zero cell occurs in those rows, mark a tick to the respective columns that contain those zeros.

(c) Examine marked columns. If any assigned zero occurs in those columns, tick the respective rows that contain those assigned zeros.

(d) Repeat this process until no more rows or columns can be marked.

(e) Draw a straight line through each marked column and each unmarked row.

Step 6: Develop the new revised opportunity cost table

(a) Among the cells not covered by any line, choose the smallest element. Call this value k.

(b) Subtract k from every element in the cell not covered by line.

(c) Add k to every element in the cell covered by the two lines, i.e. intersection of two lines.

(d) Elements in cells covered by one line remain unchanged.

Step 7: Repeat Steps 3 to 6 until an optimal solution is obtained.

Problems on Hungarian Method:

Illustration 6.1:

An institute has four employees I, II, III and IV with four jobs A, B, C and D to be performed. The following table shows the time (in days) each employee will take to perform each job. Apply Hungarian method to find how should the jobs be allocated one per employee so as to minimize the total time for completion of all jobs?

Table 6.2

Employee → / Job	I	II	III	IV
A	10	12	19	11
B	5	10	7	8
C	12	14	13	11
D	8	15	11	9

Solution:

Here the number of rows and columns are equal. So, the given assignment problem is balanced.

Step – 1: First of all find minimum element of each row in the matrix given in Table 6.2.

Table 6.3

	I	II	III	IV	Row Minimum
A	10	12	19	11	10
B	5	10	7	8	5
C	12	14	13	11	11
D	8	15	11	9	8

Subtract the minimum element of each row from all elements of that row, the matrix reduces to

Table 6.4

	I	II	III	IV
A	0	2	9	1
B	0	5	2	3
C	1	3	2	0
D	0	7	3	1

Step – 2: In reduced Table 6.4 the minimum element of columns I, II, III and IV is 0, 2, 2 and 0 respectively. Subtract these elements from all elements in their respective column, the matrix in Table 6.4 reduces to

Table 6.5

	I	II	III	IV
A	0	0	7	1
B	0	3	0	3
C	1	1	0	0
D	0	5	1	1

Step – 3: Examine all the rows, one-by-one, and find rows containing single zero element. In Table 6.5 fourth row have only one zero element in the cells (D, I). Make assignment in these cell and cross off all other zero elements in the assigned column as shown in Table 6.6.

Table 6.6

	I	II	III	IV
A	$\cancel{0}$	0	7	1
B	$\cancel{0}$	3	0	3
C	1	1	0	0
D	$\boxed{0}$	5	1	1

Now examine each column, one-by-one in Table 6.6. There is one zero in column II and column IV in the cell (A, II) and (C, IV) respectively. Make assignments in these cells and cross off all other zero elements in the corresponding rows as shown in Table 6.7.

Table 6.7

	I	II	III	IV
A	$\cancel{0}$	$\boxed{0}$	7	1
B	$\cancel{0}$	3	0	3
C	1	1	$\cancel{0}$	$\boxed{0}$
D	$\boxed{0}$	5	1	1

Still there is a cell with element 0 which is neither assigned nor crossed off. So repeat the process of assignment. Examine Table 6.7 row wise and find rows with a single unassigned and uncrossed zero element. In Table 6.7, the row B has a single unassigned and uncrossed zero element in the cell (B, III). Make an assignment here as shown in Table 6.8.

Table 6.8

	I	II	III	IV
A	$\cancel{0}$	$\boxed{0}$	7	1
B	$\cancel{0}$	3	$\boxed{0}$	3
C	1	1	$\cancel{0}$	$\boxed{0}$
D	$\boxed{0}$	5	1	1

Thus all the four assignments have been made. The optimal assignment schedule and total time is calculated below:Table 6.9

Employee	Job	Time (in days)
I	D	8
II	A	12
III	B	7
IV	C	11
		Total: 38 days

Illustration 6.2:

A wafer company producing a single product sold it through five agencies situated in different cities. There is a demand of this product in five other cities where company has no any agency. Company has to assign one city from these cities to each of the agencies in order to dispatch the product in such a way that the travelling distance is minimized. The distance between the cities having agency and cities having no agency (in Km.) is given in the following table.

Table 6.10

Cities having agency	Cities having no agency				
	a	b	c	d	e
A	160	130	175	190	200
B	135	120	130	160	175
C	140	110	155	170	185
D	50	50	80	80	110
E	55	35	70	80	105

Determine the optimum assignment schedule.

Solution:

Step – 1 and Step – 2: First of all find the minimum element in each row of the given time matrix. They are 130, 120, 110, 50 and 35. Subtracting this from all elements of the corresponding row, the resulting reduced distance matrix is shown in Table 6.11 (a). In this reduced matrix minimum element of columns are 0, 0, 10, 30 and 55. Subtract these elements from all elements in their respective column. The reduced matrix is shown in Table 6.11 (b).

Table 6.11 (a)

	A	b	c	d	E
A	30	0	45	60	70
B	15	0	10	40	55
C	30	0	45	60	75
D	0	0	30	30	60
E	20	0	35	45	70

Table 6.11 (b)

	a	b	c	d	e
A	30	0	35	30	15
B	15	0	0	10	0
C	30	0	35	30	20
D	0	0	20	0	5
E	20	0	25	15	15

Step – 3: Examine all the rows of Table 6.11 (b), one-by-one, and find rows containing single zero element. In Table 6.11 (b) first row has only one zero element in the cells (A, b). Make assignment in these cells and cross off all other zero elements in the assigned column as shown in Table 6.12. No more rows in Table 6.12 with only one uncrossed zero element.

Table 6.12

	a	b	c	d	e
A	30	$\boxed{0}$	35	30	15
B	15	$\cancel{0}$	0	10	0
C	30	$\cancel{0}$	35	30	20
D	0	$\cancel{0}$	20	0	5
E	20	$\cancel{0}$	25	15	15

Step – 4: Now examine each column, one-by-one. There is one zero in first column in the cell (D, a). Make assignment in this cell and cross off all other zeros in the assigned row as shown in Table 6.13.

Table 6.13

	a	b	c	d	e
A	30	$\boxed{0}$	35	30	15
B	15	$\cancel{0}$	$\boxed{0}$	10	$\cancel{0}$
C	30	$\cancel{0}$	35	30	20
D	$\boxed{0}$	$\cancel{0}$	20	$\cancel{0}$	5
E	20	$\cancel{0}$	25	15	15

Continue examining columns, the only zero is found in the cell (B, c). Make assignment in this cell as shown in Table 6.13. All zeros in the table are now either assigned or crossed off. This solution is not optimal because only three assignments are made.

Step – 5: Cover all zeros of the matrix by drawing minimum number of lines as explained below:

(i) Mark √ against all rows where there is no assignment (Row C and Row E)

(ii) Mark √ below the columns where × found in all the rows marked in the previous step (Column b)

(iii) Mark √ against rows where assignment found in all the columns marked in the previous step (Row A)

(iv) Repeat Steps (ii) and (iii) until no rows or columns can be marked.

(v) Draw straight lines through the unmarked rows (Row B and Row D) and marked columns (Column b) as shown in Table 6.14.

Table 6.14

	a	B	c	d	e	
A	30	$\boxed{0}$	35	30	15	√
B	15	$\cancel{0}$	$\boxed{0}$	10	$\cancel{0}$	
C	30	$\cancel{0}$	35	30	20	√
D	$\boxed{0}$	$\cancel{0}$	20	$\cancel{0}$	5	
E	20	$\cancel{0}$	25	15	15	√
		√				

Step – 6: Develop the revised matrix as described below:

Select smallest element among all uncovered elements by the lines. In Table 6.14, it is $k = 15$. Subtract this value of k from all the uncovered elements by the lines i.e. 30, 35, 30, 15, 30, 35, 30, 20, 20, 25, 15 and 15 in cells (A, a), (A, c), (A, d), (A, e), (C, a), (C, c), (C, d), (C, e), (E, a), (E, c),

(E, d) and (E, e) respectively. Add the value of k to the elements that are covered by both horizontal and vertical lines i.e. 0 and 0 in cells (B, b) and (D, b) respectively. Leave remaining cell's value unchanged. The revised matrix is shown in Table 6.15.

Table 6.15

	a	b	c	d	e
A	15	0	20	15	0
B	15	15	0	10	0
C	15	0	20	15	5
D	0	15	20	0	5
E	5	0	10	0	0

Step – 7: Repeat Step – 3 to Step – 6 until the optimal solution is obtained.

Assignments made by repeating Step – 3 and Step – 4 on Table 6.15 are shown in Table 6.16.

Table 6.16

	a	b	c	d	e
A	15	$\cancel{0}$	20	15	$\boxed{0}$
B	15	15	$\boxed{0}$	10	$\cancel{0}$
C	15	$\boxed{0}$	20	15	5
D	$\boxed{0}$	15	20	$\cancel{0}$	5
E	5	$\cancel{0}$	10	$\boxed{0}$	$\cancel{0}$

In Table 6.16 there are five assignments which is equal to the number of rows. So the optimum solution is reached.

The pattern of assignments with their respective distance is given in Table 6.17.

Table 6.17

Cities having agency	Cities having no agency	Distance (in Km)
A	E	200
B	C	130
C	B	110
D	A	50
E	D	80
		Total: 570

6.5 OPTIMAL SOLUTION WITH TIE FOR ASSIGNMENT

Illustration 6.3:

Find the optimum assignment from the following data matrix.

Table 6.18

	I	II	III	IV	V
A	5	0	3	2	6
B	0	0	5	4	7
C	0	3	0	4	0
D	0	1	0	3	0
E	6	5	0	0	0

Solution:

Step – 1 and Step – 2: Observe that number of rows and number of columns in the given data matrix are same and all the rows and columns of given data matrix have at least one zero. Hence, the given table itself is an opportunity cost table.

Step – 3: Examine all the rows of Table 6.18, one-by-one, and find rows containing single zero element. In Table 6.18 first row has only one zero element in the cell (A, II). Make assignment in this cell and cross off all other zero elements in the assigned column as shown in Table 6.19.

Table 6.19

	I	II	III	IV	V
A	5	$\boxed{0}$	3	2	6
B	0	$\cancel{0}$	5	4	7
C	0	3	0	4	0
D	0	1	0	3	0
E	6	5	0	0	0

Continuing this process, row B have only one uncrossed and unassigned zero element in the cell(B, I). So, make next assignment in the cell (B, I) and cross off all other zero elements in the assigned column. Observe that all the remaining rows don’t have single unassigned and uncrossed zero element. So, now by examining column wise the next assignment can be done in the cell (E, IV) as shown in Table 6.20

Table 6.20

	I	II	III	IV	V
A	5	$\boxed{0}$	3	2	6
B	$\boxed{0}$	$\cancel{0}$	5	4	7
C	$\cancel{0}$	3	0	4	0
D	$\cancel{0}$	1	0	3	0
E	6	5	$\cancel{0}$	$\boxed{0}$	$\cancel{0}$

Again examining Table 6.20 row wise there is no possibility of assignment but there are four cells (C, III), (C, V), (D, III) and (D, V) with uncrossed and unassigned zero element. This situation indicates tie for assignment in cells (C, III), (D, V) and (C, V), (D, III). If we select the cell (C, III) for assignment, the next assignment we can do in the cell (D, V). Alternatively, by choosing the cell (C, V) for assignment, we have to make assignment in the cell (D, III) as shown in Table 6.21 (a) and Table 6.21 (b).

Table 6.21 (a)

	I	II	III	IV	V
A	5	$\boxed{0}$	3	2	6
B	$\boxed{0}$	$\cancel{0}$	5	4	7
C	$\cancel{0}$	3	$\boxed{0}$	4	$\cancel{0}$
D	$\cancel{0}$	1	$\cancel{0}$	3	$\boxed{0}$
E	6	5	$\cancel{0}$	$\boxed{0}$	$\cancel{0}$

Table 6.21 (b)

	I	II	III	IV	V
A	5	$\boxed{0}$	3	2	6
B	$\boxed{0}$	$\cancel{0}$	5	4	7
C	$\cancel{0}$	3	$\cancel{0}$	4	$\boxed{0}$
D	$\cancel{0}$	1	$\boxed{0}$	3	$\cancel{0}$
E	6	5	$\cancel{0}$	$\boxed{0}$	$\cancel{0}$

The pattern of these two alternative optimal assignments is given in Table 6.22 (a) and Table 6.22 (b).

Table 6.22 (a)

A	**II**
B	I
C	III
D	V
E	IV

Table 6.22 (b)

A	**II**
B	I
C	V
D	III
E	IV

Illustration 6.4:

Five employees are available to do five different jobs. The time (in hours) that each employee takes to do each job is given in the following table:

Table 6.23

	Job				
Employee	I	II	III	IV	V
A	2	9	2	7	1
B	6	8	7	6	1
C	4	6	5	3	1
D	4	2	7	3	1
E	5	3	9	5	1

Find out how employees should be assigned the jobs in way that will minimize the total time taken.

Solution:

Step – 1 and Step – 2: Observe that the minimum element in each row of the given time matrix is 1. Subtracting this from all elements of each row, the resulting reduced time matrix is shown in Table 6.24 (a). In this reduced matrix minimum element of columns are 1, 1, 1, 2 and 0. Subtract these elements from all elements in their respective column. The reduced matrix is shown in Table 6.24 (b).

Table 6.24 (a)

	I	II	III	IV	V
A	1	8	1	6	0
B	5	7	6	5	0
C	3	5	4	2	0
D	3	1	6	2	0
E	4	2	8	4	0

Table 6.24 (b)

	I	II	III	IV	V
A	0	7	0	4	0
B	4	6	5	3	0
C	2	4	3	0	0
D	2	0	5	0	0
E	3	1	7	2	0

Step – 3: Examine all the rows of Table 6.24 (b), one-by-one, and find rows containing single zero element. In Table 6.24 (b) second row has only one zero element in the cells (B, V). Make assignment in these cell and cross off all other zero elements in the assigned column as shown in Table 6.25.

Table 6.25

	I	II	III	IV	V
A	0	7	0	4	$\cancel{0}$
B	4	6	5	3	$\boxed{0}$
C	2	4	3	0	$\cancel{0}$
D	2	0	5	0	$\cancel{0}$
E	3	1	7	2	$\cancel{0}$

Continuing this process, rows C and D have only one zero element in the cells (C, IV) and (D, II). Make assignment in these cells, and cross off all other zeros in the assigned column as shown in Table 6.26.

Step – 4: Now examine each column, one-by-one. There is one zero in column I in the cell (A, I). Make assignment in this cell and cross off all other zeros in the assigned row as shown in Table 6.26. All zeros in the table are now either assigned or crossed off. This solution is not optimal because only four assignments are made.

Table 6.26

	I	II	III	IV	V
A	$\boxed{0}$	7	$\cancel{0}$	4	$\cancel{0}$
B	4	6	5	3	$\boxed{0}$
C	2	4	3	$\boxed{0}$	$\cancel{0}$
D	2	$\boxed{0}$	5	$\cancel{0}$	$\cancel{0}$
E	3	1	7	2	$\cancel{0}$

Step – 5: Cover all zeros of the matrix by drawing minimum number of lines as explained below:

(i) Mark √ against all rows where there is no assignment (Row E)

(ii) Mark √ below the columns where × found in all the rows marked in the previous step (Column V)

(iii) Mark √ against rows where assignment found in all the columns marked in the previous step (Row B)

(iv) Repeat Steps (ii) and (iii) until no rows or columns can be marked.

(v) Draw straight lines through the unmarked rows (Row) and marked columns as shown in Table 6.27.

Table 6.27

	I	II	III	IV	V	
A	$\boxed{0}$	7	$\cancel{0}$	4	$\cancel{0}$	
B	4	6	5	3	$\boxed{0}$	√
C	2	4	3	$\boxed{0}$	$\cancel{0}$	
D	2	$\boxed{0}$	5	$\cancel{0}$	$\cancel{0}$	
E	3	1	7	2	$\cancel{0}$	√
					√	

Step – 6: Develop the revised matrix as described below:

Select smallest element among all uncovered elements by the lines. In Table 6.27, it is $k = 1$ in the cell (E, II). Subtract this value of k from all the uncovered elements by the lines i.e. 4, 6, 5, 3, 3, 1, 7 and 2 in cells (B, I), (B, II), (B, III), (B, IV), (E, I), (E, II), (E, III) and (E, IV) respectively. Add the value of k to the elements that are covered by both horizontal and vertical lines i.e. 0, 0 and 0 in cells (A, V), (C, V) and (E,V). The revised matrix is shown in Table 6.28.

Table 6.28

	I	II	III	IV	V
A	0	7	0	4	1
B	3	5	4	2	0
C	2	4	3	0	1
D	2	0	5	0	1
E	2	0	6	1	0

Step – 7: Repeat Step – 3 to Step – 6 until the optimal solution is obtained.

Assignments made by repeating Step – 3 and Step – 4 on Table 6.28 are shown in Table 6.29.

Table 6.29

	I	II	III	IV	V
A	$\boxed{0}$	7	~~0~~	4	1
B	3	5	4	2	$\boxed{0}$
C	2	4	3	$\boxed{0}$	1
D	2	$\boxed{0}$	5	~~0~~	1
E	2	~~0~~	6	1	~~0~~

Covering of all zeros of Table 6.29 by minimum number of lines is shown in Table 6.30.

Table 6.30

	I	II	III	IV	V	
A	$\boxed{0}$	7	~~0~~	4	1	
B	3	5	4	2	$\boxed{0}$	√
C	2	4	3	$\boxed{0}$	1	√
D	2	$\boxed{0}$	5	~~0~~	1	√
E	2	~~0~~	6	1	~~0~~	√
		√		√	√	

The smallest element among all uncovered elements in Table 6.30 is 2. So by repeating Step – 6,the revised matrix can be written as shown in Table 6.31.

Table 6.31

	I	II	III	IV	V
A	0	9	0	6	3
B	1	5	2	2	0
C	0	4	1	0	1
D	0	0	3	0	1
E	0	0	4	1	0

Again repeat Step – 3 and Step – 4 on Table 6.31 for making assignment. There is a one zero in row B in the cell (B, V). Make assignment in this cell and cross off all other zeros in the column

V. Now examine columns one-by-one, column III has one zero in the cell (A, III). Make assignment in this cell and cross off other zeros in the row A. These two assignments are shown in Table 6.32. Again examining one-by-one rows of Table 6.32, row C has two zeros in the cell (C, I) and (C, IV). We can choose any one of these cell for making assignment. Let us assign in the cell (C, I) and cross off all other zeros of the row C and column I. Continue the process of assignment as shown in Table 6.33.

Table 6.32

	I	II	III	IV	V
A	$\cancel{0}$	9	$\boxed{0}$	6	3
B	1	5	2	2	$\boxed{0}$
C	0	4	1	0	1
D	0	0	3	0	1
E	0	0	4	1	$\cancel{0}$

Table 6.33

	I	II	III	IV	V
A	$\cancel{0}$	9	$\boxed{0}$	6	3
B	1	5	2	2	$\boxed{0}$
C	$\boxed{0}$	4	1	$\cancel{0}$	1
D	$\cancel{0}$	$\cancel{0}$	3	$\boxed{0}$	1
E	$\cancel{0}$	$\boxed{0}$	4	1	$\cancel{0}$

In Table 6.33 there are five assignments which is equal to the number of rows. So the optimum solution is reached.

The pattern of assignments with their respective time is given in Table 6.34.

Table 6.34

Employee	Job	Time (in hours)
A	III	2
B	V	1
C	I	4
D	IV	3
E	II	3
		Total: 13

Alternatively, if we make assignment in the cell (C, IV) instead of the cell (C, I), there are two possibilities of assignments as shown in Table 6.35 (a) and 6.35 (b).

Table 6.35 (a)

	I	II	III	IV	V
A	$\cancel{0}$	9	$\boxed{0}$	6	3
B	1	5	2	2	$\boxed{0}$
C	$\cancel{0}$	4	1	$\boxed{0}$	1
D	$\boxed{0}$	$\cancel{0}$	3	$\cancel{0}$	1
E	$\cancel{0}$	0	4	1	$\cancel{0}$

Table 6.35 (b)

	I	II	III	IV	V
A	$\cancel{0}$	9	$\boxed{0}$	6	3
B	1	5	2	2	$\boxed{0}$
C	$\cancel{0}$	4	1	$\boxed{0}$	1
D	$\cancel{0}$	$\boxed{0}$	3	$\cancel{0}$	1
E	0	$\cancel{0}$	4	1	$\cancel{0}$

The pattern of these two alternative optimal assignments of jobs among employees with their respective time (in hours) is given in Table 6.36 (a) and Table 6.36 (b).

Table 6.36 (a)

Employee	Job	Time (in hours)
A	III	2
B	V	1
C	IV	3
D	I	4
E	II	3
		Total: 13

Table 6.36 (b)

Employee	Job	Time (in hours)
A	III	2
B	V	1
C	IV	3
D	II	2
E	I	5
		Total: 13

6.6 PROFIT ASSIGNMENT (ASSIGNMENT WITH MAXIMIZATION)

Illustration 6.5:

In a reputed mall there are five counters and five employees. The profit on each counter depends on the efficiency of the employee. Following is a matrix shows the monthly profit (in '000 Rs.) on each counter according to the employee assigned to it.

Table 6.37

Employee	Counter				
	I	II	III	IV	V
A	30	37	40	27	40
B	40	24	26	21	36
C	40	32	33	30	35
D	25	38	40	36	36
E	29	60	41	34	39

How the counters should be assigned to employees so as to maximize the total profit.

Solution:

The given maximization problem can be converted into a minimization problem by subtracting all the elements of the Table 6.37 from the largest element i.e. 60. The new data so obtained is given in Table 6.38.

Table 6.38

Employee	Counter				
	I	II	III	IV	V
A	30	23	20	33	20
B	20	36	34	39	24
C	20	28	27	30	25
D	35	22	20	24	24
E	31	0	19	26	21

Subtracting row minimum from all elements in respective rows and then column minimum from all elements in their respective columns, the reduced matrix so obtained is shown in Table 6.39.

Table 6.39

	I	II	III	IV	V
A	10	3	0	9	0
B	0	16	14	15	4
C	0	8	7	6	5
D	15	2	0	0	4
E	31	0	19	22	21

Make assignment in Table 6.39 by applying Hungarian method, as shown in Table 6.40.

Table 6.40

	I	II	III	IV	V
A	10	3	~~0~~	9	$\boxed{0}$
B	$\boxed{0}$	16	14	15	4
C	~~0~~	8	7	6	5
D	15	2	~~0~~	$\boxed{0}$	4
E	31	$\boxed{0}$	19	22	21

The solution shown in Table 6.40 is not optimal since only four assignments are made. Cover all zeros of Table 6.40 with the minimum number of lines as shown in Table 6.41.

Table 6.41

	I	II	III	IV	V	
A	10	3	~~0~~	9	$\boxed{0}$	
B	$\boxed{0}$	16	14	15	4	√
C	~~0~~	8	7	6	5	√
D	15	2	~~0~~	$\boxed{0}$	4	
E	31	$\boxed{0}$	19	22	21	
	√					

Revised matrix as shown in Table 6.42 is obtained by subtracting minimum element among all uncovered elements from all uncovered elements and adding it to the elements that are covered by both horizontal and vertical lines.

Table 6.42

	I	II	III	IV	V
A	14	3	$\boxed{0}$	9	$\cancel{0}$
B	$\cancel{0}$	12	10	11	$\boxed{0}$
C	$\boxed{0}$	4	3	2	1
D	19	2	$\cancel{0}$	$\boxed{0}$	4
E	35	$\boxed{0}$	19	22	21

This solution is optimum as five assignments done. The pattern of this optimal assignments of counters among employee with their respective profit (in '000 Rs) is given in Table 6.43.

Table 6.43

Employee	Counter	Profit (in '000 Rs.)
A	III	40
B	V	36
C	I	40
D	IV	36
E	II	60
		Total: 212

6.7 UNBALANCED ASSIGNMENT PROBLEM

The Hungarian method can be applied to an assignment problem for its solution only if the number of rows and columns in the assignment matrix are same, that is, the assignment matrix is a square matrix. When the given assignment matrix is not a square matrix, the assignment problem is called an unbalanced assignment problem. In such a case before applying Hungarian method, dummy row(s) or column(s) should be added in the matrix with zeros as the cost elements in order to make it a square matrix.

Illustration 6.6:

There are four tasks to perform in a store and three employees. The employees differ in efficiency. The estimates of the time (in hours), each employee would take to perform, is given in the following table.

Table 6.44

Tasks	Employee		
	I	II	III
A	9	26	15
B	13	27	6
C	35	20	15
D	18	30	20

How should the tasks be allocated, one to each employee, so as to minimize the total time?

Solution:

Since the time matrix given in Table 6.44 is not square, we add one dummy column (Employee IV) with all zero elements in that column as shown in Table 6.45.

Table 6.45

	I	II	III	IV
A	9	26	15	0
B	13	27	6	0
C	35	20	15	0
D	18	30	20	0

Subtracting column minimum from all elements in their respective columns, the reduced matrix so obtained is shown in Table 6.46.

Table 6.46

	I	II	III	IV
A	0	6	9	0
B	4	7	0	0
C	26	0	9	0
D	9	10	14	0

Perform the Hungarian method on Table 6.46 to make assignments as shown in Table 6.47.

Table 6.47

	I	II	III	IV
A	$\boxed{0}$	6	9	~~0~~
B	4	7	$\boxed{0}$	~~0~~
C	26	$\boxed{0}$	9	~~0~~
D	9	10	14	$\boxed{0}$

Since the number of assignments in Table 6.47 is equal to the number of rows or columns, the solution is optimal. The total minimum time (in hours) and optimal assignments made are shown in Table 6.48.

Table 6.48

Task	Employee	Tine (in hours)
A	I	9
B	III	6
C	II	20
D	IV	0
		Total: 35

❖ **CHECK YOUR PROGRESS**

- **Answer the following multiple choice questions.**

Que. 1 If the number of rows and columns in an assignment problem are not equal then it iscalled______________problem.

(a) prohibited (b) infeasible

(c) unbounded (d) unbalanced

Que. 2 Which of the following method is a method of solution of assignment problem?

(a) North–West corner (b) least cost method

(c) Hungarian (d) VAM

Que. 3 The extra row or column which is added to balance an assignment problem is called _________row or column.

(a) regret (b) dummy

(c) extra (d) none of these

Que. 4 An assignment problem is considered as a particular case of a Transportation problembecause

(a) the number of rows and columns are equal

(b) all $x_{\mathrm{ij}} = 0$

(c) all rim conditions are 1

(d) all of the above

Que. 5 An optimal of an assignment problem can be obtained only if

(a) each row and column has only one zero element

(b) each row and column has at least one zero element

(c) the data are arranged in a square matrix

(d) none of the above

Que. 6 In an assignment problem

(a) one agent can do parts of several tasks

(b) one task can be done by several agents

(c) each agent is assigned to its own best task

(d) none of the above

- **Answer the following questions in brief.**

Que. 1 What is an assignment problem?

Que. 2 What is the difference between transportation problem and assignment problem?

Que. 3 State the mathematical formulation for assignment problem.

Que. 4 Write the steps (in brief) for solving an assignment problem by Hungarian method.

Que. 5 Write name of different methods to obtain the solution of an assignment problem.

Que. 6 According to Hungarian method, how to revise the matrix after all zeros of the matrix is covered by minimum number of lines?

Que. 7 What do you mean by unbalanced assignment problem?

- **Answer the following questions in detail.**

Que. 1 Write the steps in detail for solving an assignment problem by Hungarian method.

Que. 2 A departmental has five employees with five jobs to be performed. The time (in hours) each men will take to perform each job is given in the effectiveness matrix. How should the jobs be allocated, one per employee, so as to minimize the total man--hours.

Employees → / Jobs	1	2	3	4	5
A	10	5	13	15	16
B	3	9	18	13	6
C	10	7	2	2	2
D	7	11	9	7	12
E	7	9	10	4	12

Que. 3 Find the assignment of salesman to various districts which will result minimum cost.

District → / Salesman	1	2	3	4
A	16	10	14	11
B	14	11	15	15
C	15	15	13	12
D	13	12	14	15

Que. 4 A city corporation has decided to carry out road repairs on main four arteries of the city. The government has agreed to make a special grant of Rs 50 lakh towards the cost with a condition that repairs are done at the lowest cost and quickest time. If the conditions warrant, a supplementary token grant will also be considered favorably. The corporation has floated tenders and five contractors have sent in their bids. In order to expedite work, one road will be awarded to only one contractor.

Cost of Repairs (Rs. In lakhs)

Road → / Contractors	R1	R2	R3	R4	R5
C1	9	14	19	15	13

C2	7	17	20	19	18
C3	9	18	21	18	17
C4	10	12	18	19	18
C5	10	15	21	16	15

Find the best way of assigning the repair work to the contractors and the costs. If it is necessary to seek supplementary grants, what should be the amount sought?

Que. 5 Solve the following assignment problem so as to minimize the time (in days) required to complete all the task.

Task → / Person	1	2	3	4	5
A	6	5	8	11	16
B	1	13	16	1	10
C	16	11	8	8	8
D	9	14	12	10	16

UNIT - 7 CPM & PERT

7.1 INTRODUCTION

A Project such as setting up of a new silk mills, research and development, location of the facility, development of a new silk mill, marketing of a product, etc. is a combination of interrelated activities (tasks) and the same is required to executed in a certain order before the entire task can be completed. These tasks are highly interrelated and possess a logical sequence. The formation of the logical sequence, in such a way that same activities can not start until some others activity/s is/are completed. In the language of quantitative management, an activity in a project usually viewed as job and it requires some amount of resources of the organisation for its completion.

A Network is a logical arrangement of the activities to be carried out within the minimum stipulated time period without compromising the quality and with minimum utilisation of resources of the organisation. Several critical terminologies should be understood before proceeding with the understanding of Network Analysis.

Network : It is diagrammatic representation of logical sequence of all the essential activities of the project.

Project : It is a predefined task/goal to be achieved. For example, establishment of Silk mills or starting an online shop or opening a Bank branch or starting a pizza delivery centre.

Activity : It is the smallest component of the project.

Event : The starting and end points of an activity are called events (or nodes). It is just a point in time and does not requires any resources. The starting and ending of an activity are known as tail event and head event respectively. Event is generally represented by a numbered circle.

Predecessor activity : An activity that must be accomplished before the start of next activity. For establishment of silk mills, construction of building is a predecessor activity to purchasing machines or for opening a Bank branch, construction of building is preceding activity to buying computers.

Successor activity : An activity that must be accomplished after the end of earlier activity. For establishment of silk mills, purchasing machines is a successor activity to construction of building or for opening a Bank branch, buying computers is a successor activity to construction of building.

Concurrent activities : These are activities that must be accomplished simultaneously with each other. The concurrent activities show its importance in network by reducing overall project completion time due to its concurrent nature.

Dummy activity : It is not actual activity and does not occupy resources. However, it is a part of network to accommodate the timings of the other activities. It indicates only precedents relationship.

Path : A path is a set of nodes connected by lines that starts at the starting activity and ends at terminal node of the given network.

Project : A project is a combination of several interrelated activities which must be executed in a certain order for its completion.

A few important questions that leads to generation of understanding about Network analysis are as follows:

i. What will be the expected time of project completion?
ii. What is the effect of delay of any activity on the overall completion of project?
iii. How to reduce the time to perform certain activities in case of availability of additional funds?
iv. What is the probability of completion of project in time?

A network is a graphical representation of activities and it consists of certain configuration of arrows and nodes for showing the logical sequence of various activities to be performed to achieve the project objectives. Before few years the planning tool such as Gantt chart used to specify the start and finish time for each activity on horizontal scale of time. There were two major limitations of this method; first, the method was not able to capture such activities which are non-interdepend on each other and second, the method was not able to capture complex relationship of dependency among several activities. In present days, the usage of formal network techniques such as PERT (Programme Evaluation and Review Technique), CPM (Critical Path Method), PEP (Programme Evaluation Procedure), LCES (Least Cost

Estimating and Scheduling), SCANS (Scheduling and Control by Automated Network System), Maximum flow method, Shortest route method, etc.
During further part of the chapter, we shall discuss about the following three methods:

a) Critical Path Method (CPM) and
b) Programme Evaluation and Review Technique (PERT)
c) Maximum flow method
d) Shortest route method

7.2 CRITICAL PATH METHOD (CPM)

Critical Path Method (CPM) and Project Evaluation and Review Techniques (PERT) are two very popular networking methods. The two techniques were developed by two different set of people with two different objectives during 1956- 1958. Critical Path Method (CPM) was first developed by E.Idu Pout de Nemours & company as an application to construction projects and was later extended to a more advanced status by Mauchly associates. While, Project Evaluation and Review Techniques (PERT), on the other hand, was developed for the U.S. Navy by a consulting firm for scheduling the research and development activities for the Polaris missile program.

Both CPM and PERT are basically time –oriented methods in the sense that they both lead to the determination of a time schedule. Although the two methods were developed independently, they are strikingly similar. Perhaps the most important difference is that originally the time estimates for the activities were assumed deterministic in CPM and probabilistic in PERT. Despite of the same there are several comparisons among these methods and the same will be explained post PERT topic.

With the view to understand CPM, let us consider a real life example. A construction company has identified seven major activities at the time to complete house construction project. As we understand earlier, the starting of one activity would be post completion of the other activity. Therefore, the construction company has also identified the essential activity to be completed before the start of the next activity and it is given in table. The time period for each activity are estimated with the help of experts and the same are tabulated as under:

Activity ID	Description of Activity	Time required (in no. of weeks)	Predecessor activities
A	Design House	2	-
B	Obtain Fund	3	-
C	Finalise locality	4	-
D	Land planning	1	A
E	Construction Material Order	2	B
F	Decorative material order	5	B
G	Decide doors, windows, and furniture design and locations	7	C
H	Construct house	2	D, E
I	Furniture work	3	F, G
J	Opening ceremony preparation	1	H, I

For the above project, let us understand the step by step process to apply CPM and construction of network for it.

Step 1: Start with first activity as node (Node represents as a circle in diagram and shows beginning/ending of the activity). Here all three activities A, B and C are concurrent activities and they start simultaneously. Nord/Event 1 (Circle 1) is considered as a starting point for Activity A, B and C (refer figure 1).

Step 2: Design node 2 for activity D. By referring to above table, we can see that activity D has predecessor activity A. It means till activity A is not completed, activity D cannot be started. Therefore, activity D begins from the node 2 which is ending point of activity A (refer figure 2).

Step 3: Similarly, activity E and F both have predecessor activity B and therefore from node 3 (ending point of activity B), both activity E and F start (refer figure 3).

Step 4: following the same method, activity G can be plotted from node 4. (refer figure 4)

Step 5: Both activity D and E end at event 5 (node 5) and the same node is a starting point for activity H, as activity H required predecessor D and E (refer figure 6). Similar, activity F and G constructs node 6 as a start for activity I. and both H and I connect at node 7 to start activity J (refer figure 7).

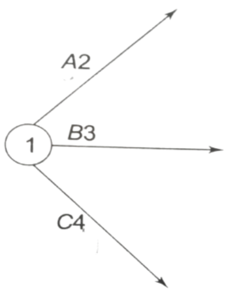


Figure 1

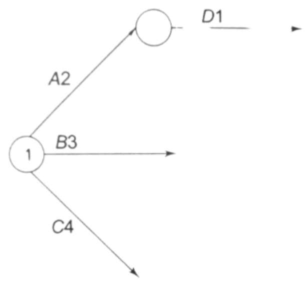


Figure 2

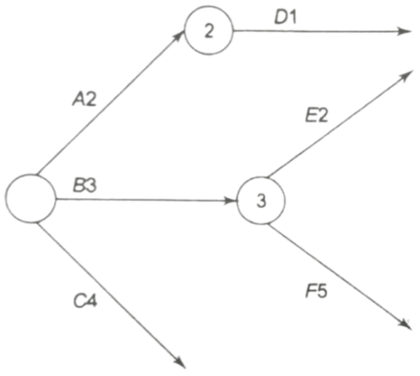


Figure 3

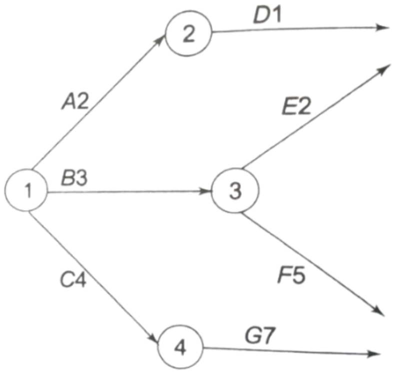


Figure 4

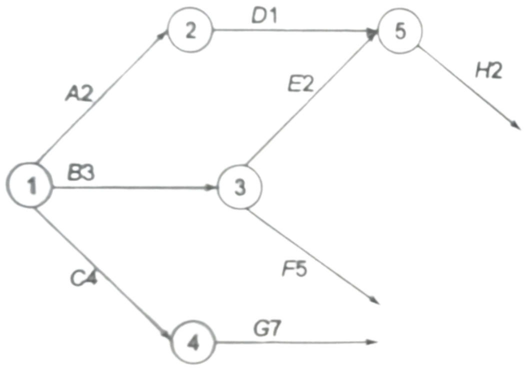


Figure 5

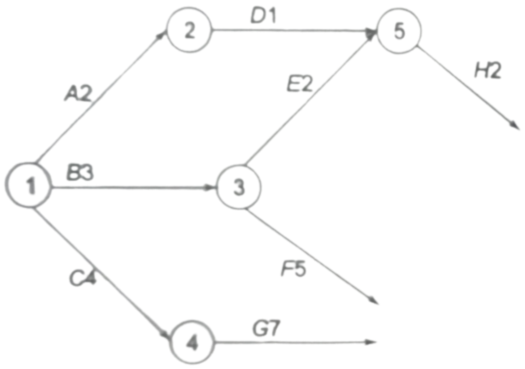


Figure 6

Critical path is the longest time require to finish the project. Once the network diagram is ready there are two approaches to find the critical path.

1) Trial and Error Method (Assessment of all possible path)
2) Forward and backward move Method (Systematic approach)

Let us understand both the methods one by one.

1) Trial and Error Method : Under this method, all the possible paths are considered and calculated for total time required to move from first node to the last node. If we refer figure 6, there are four possible path to move from node 1 to node 8; 1-2-7-8, 1-3-5-8, 1-3-6-7-8, 1-4-6-7-8. Let us calculate the duration (total time required) for each path

Path	Time for each activity	Duration (Total time)
1-2-5-7-8 (ADHJ)	2+1+2+1	6
1-3-5-7-8 (BEHJ)	3+2+2+1	8
1-3-6-7-8	3+5+3+1	12
1-4-6-7-8 (CGIJ)	4+7+3+1	15

Considering the above table, find the path that has highest duration and it is called Critical Path. Therefore, for the given example Path 1-4-6-7-8 (CGIJ) is a critical path. It looks like ad-hoc method and need to check each and every path. Just consider the following network and try to understand the different paths and duration for each of them (figure 7). Isn't it really difficult or chances of missing any path for calculation of duration?

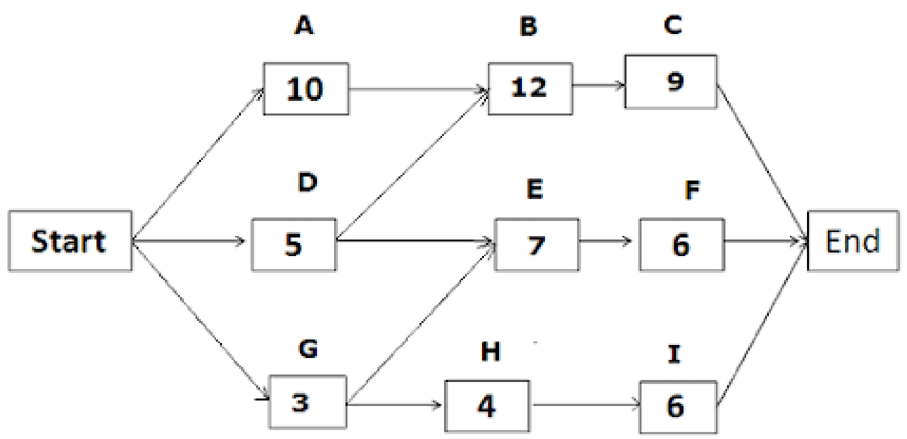


Figure 7

2) Forward and Backward Move method : As the trial and error method pose greater difficulties in identifying the path with longest time duration, there was a constant strive to identify some systematic and scientific method to identify critical path. Forward and Backward move method is the solution for the same. Let us understand the forward and backward move method to identify critical path.

Forward move (Earliest start and finish time): Let us calculate earliest start and finish time from starting to ending of the project and for each activity.

Step 1: Identify the forward move from node 1 to 2, 1 to 3 and 1 to 4 (Consider figure 6). From node 1 to 2 there is an activity A with time period of 2 week. Therefore, the path A to 2 will be finished earliest on 2nd unit of time, if started at 0th week. Similarly, for node 1 to 3, if starting time is 0th week the earliest finish time will be 3rd week and similarly for node 1 to 4, it is 4th week (refer figure 8).

Step 2 : Repeat step 1 for node 2 to 5 and 3 to 5. It is found that for path 2 to 5, the earliest stating is possible with 2nd week and earliest finish is possible at 3rd week. However, for node 3 to 5, the earliest stating time is 3rd week and earliest finish time is 5th week (refer figure 9). Similarly count the same for node 3 to 6 and node 4 to 6 to arrive earliest start and earliest finish time.

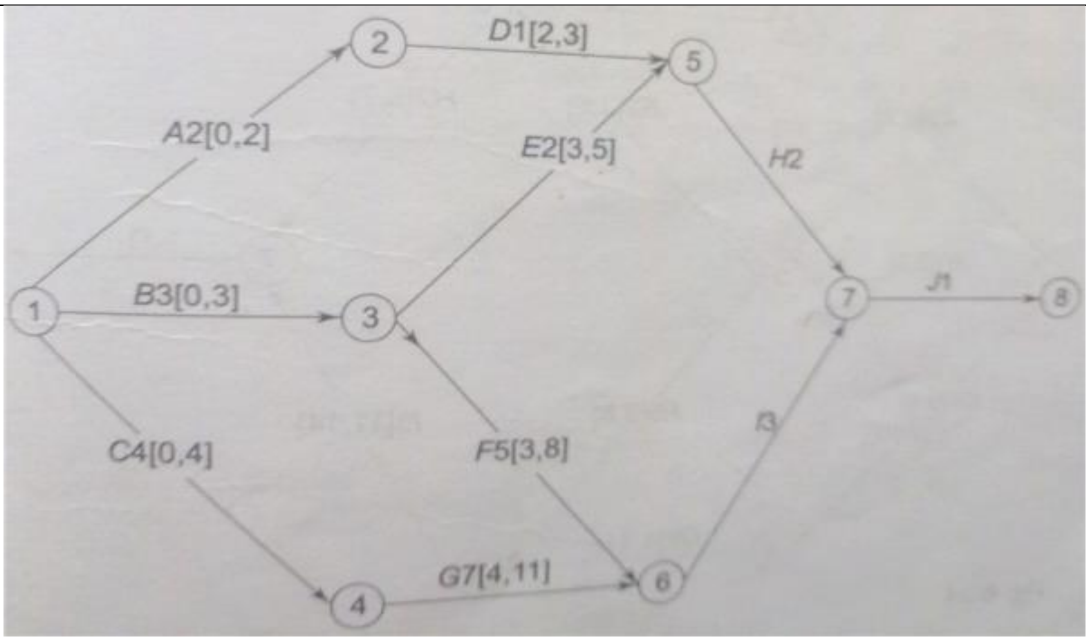


Figure 8

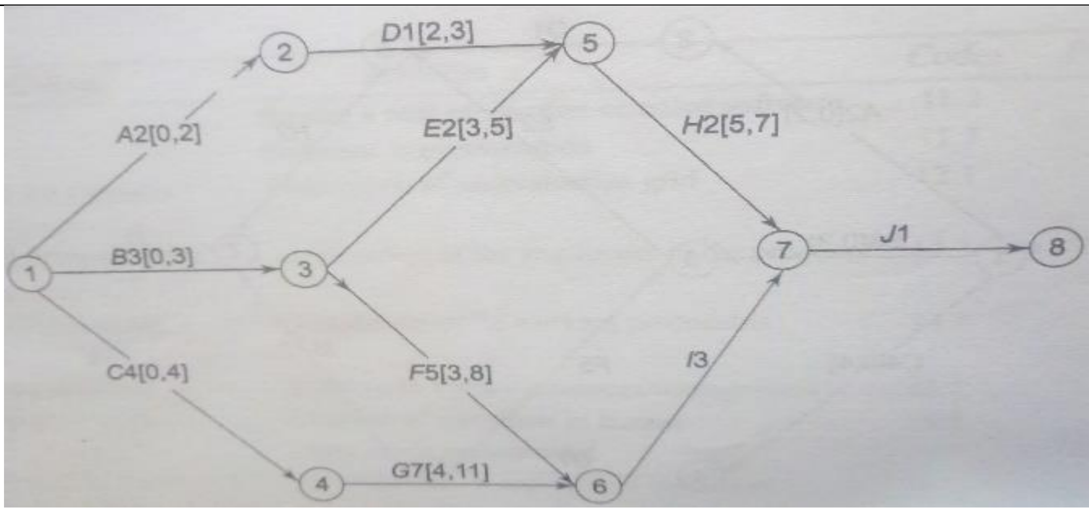


Figure 9

Step 3 : This stage is crucial as both node 5 and 6 meet at node 7. For node 5, earliest finish time is 5th week (As unless activity E is completed, activity 5 cannot be started). Similarly, for node 6, earliest finish time is 11th week (As unless activity G is completed, node 6 cannot be started). Therefore, for node 7, earliest starting time will be 11th week and not before that (refer figure 10).

Step 4 : Path 7 to 8 can be completed in 1 unit of time therefore, the earliest start time for this path is 11th week and earliest finish time is 12th week (refer figure 11).

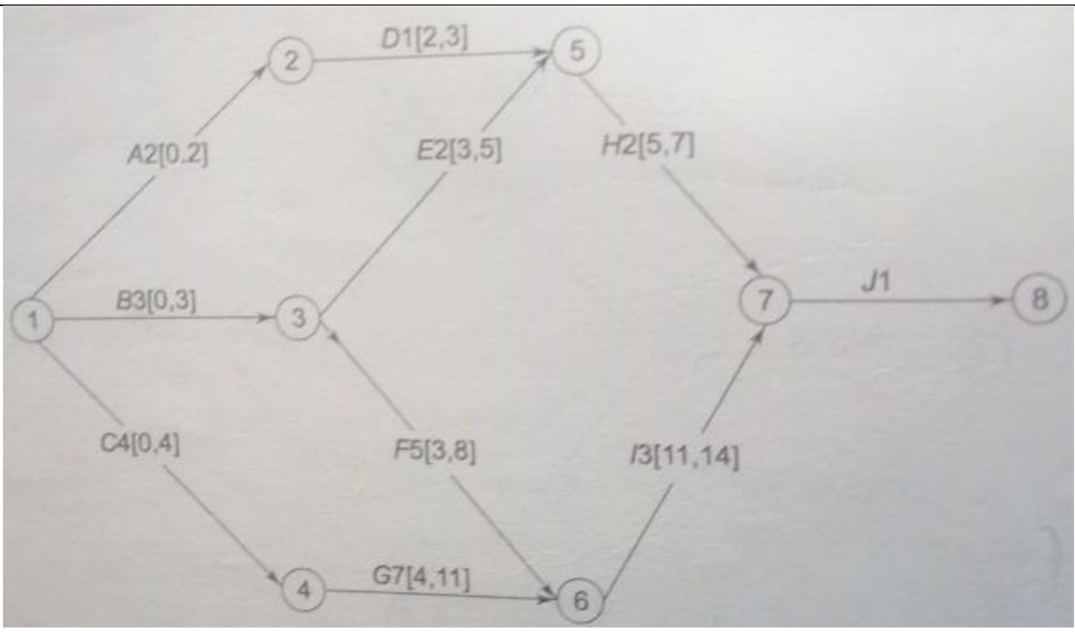


Figure 10

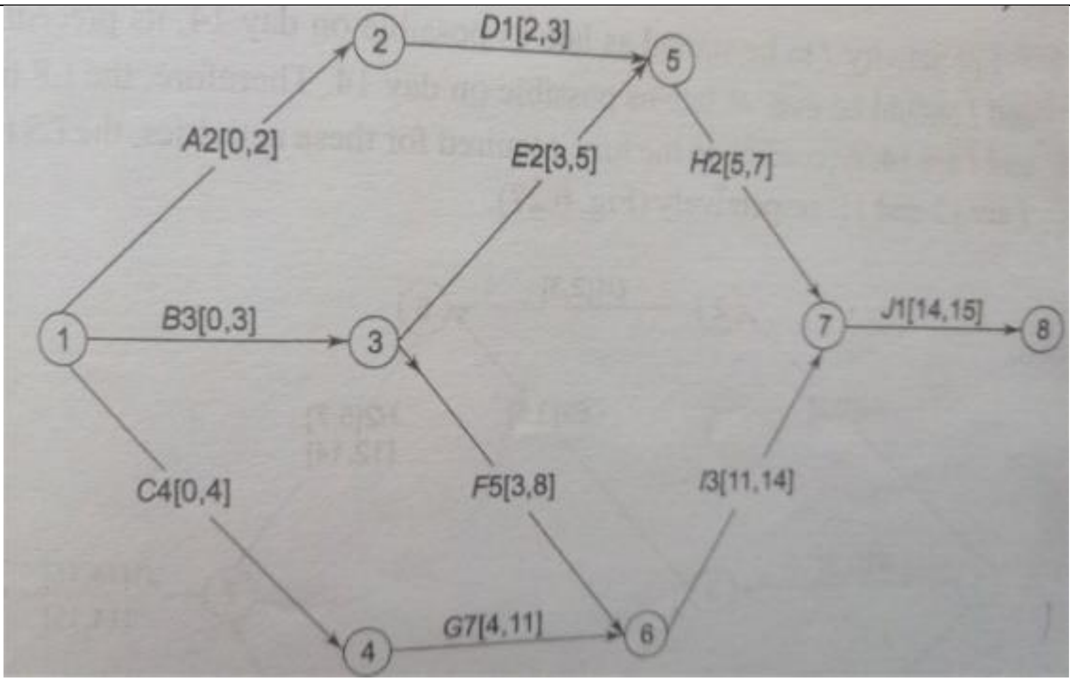


Figure 11

Backward move (Latest start and finish time): Let us calculate latest start and finish time from ending to starting of the project and for each activity.

Step 1 : In our example, the project has to be over on 15th unit of time. Therefore, for the last activity J of the project, we keep latest finish time as 15. Activity J requires 1 week and hence its latest start time will be 15-1 = 14th week (refer figure 12).

Step 2 : For activity J to be started as late possible on week 14, its preceding activities H and I should over as late as possible on day 14. Therefore, the latest finish for H and I will be 14th week and latest start will be 12th and 11th week respectively for H and I activities (refer figure 13).

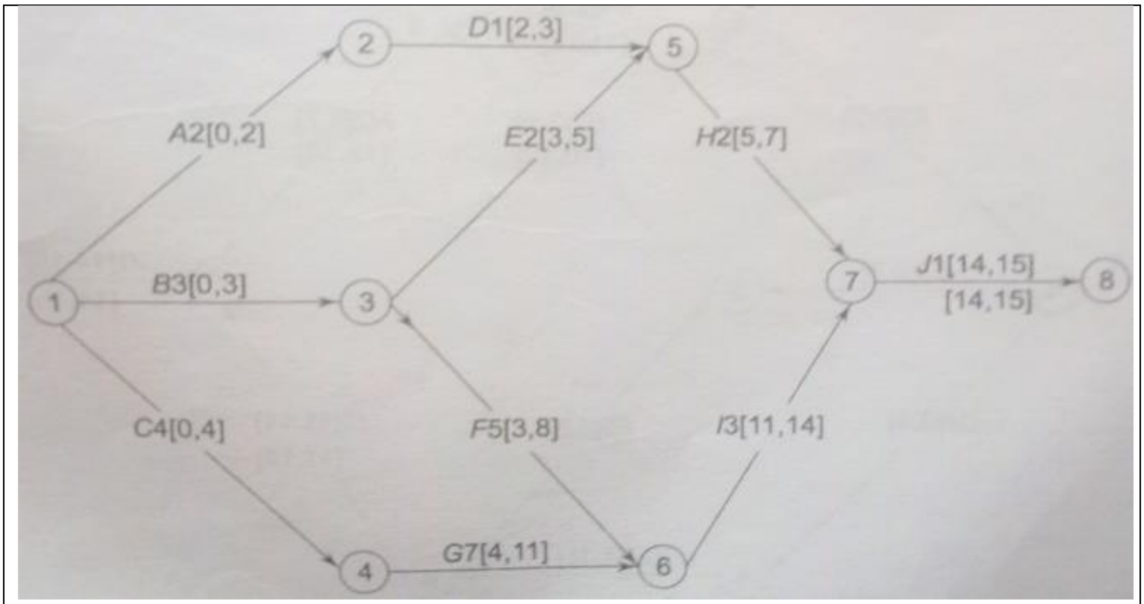


Figure 12

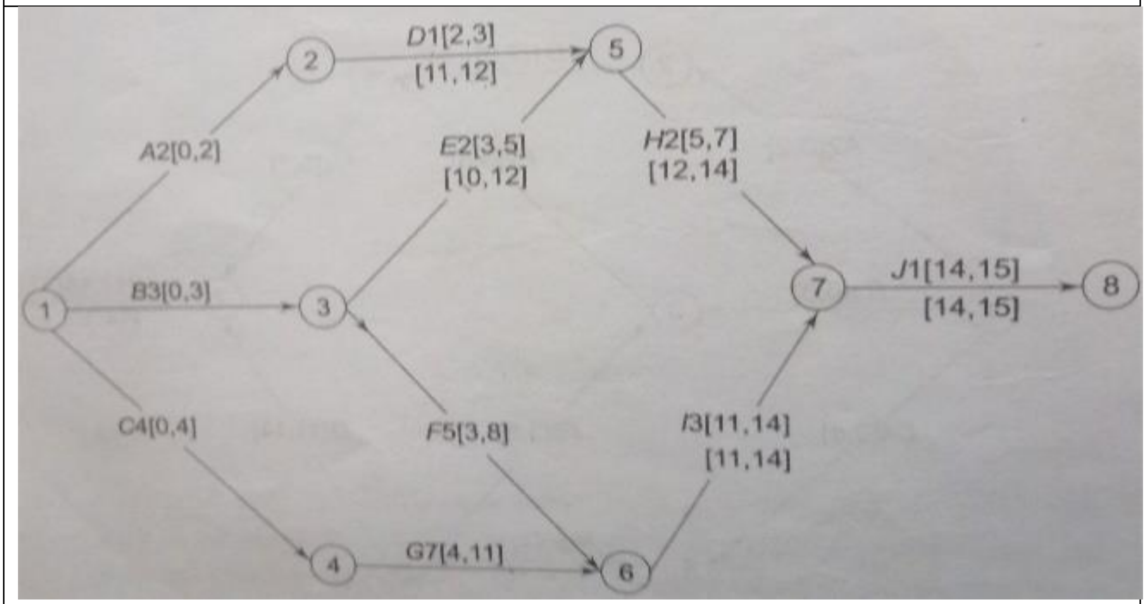


Figure 13

Step 3 : Similarly, we need to calculate latest finish and start time for each activities and the same are considered in following charts (refer figure 14, 15, and 16).

Step 4 : Once both forward and backward moves are completed, the time is to identify a path having same timings for both forward and backward moves. It is clear that there will be only one path that has same value in both the cases of forward and backward moves. This path (from first node to the last) is referred as Critical Path (refer figure 17).

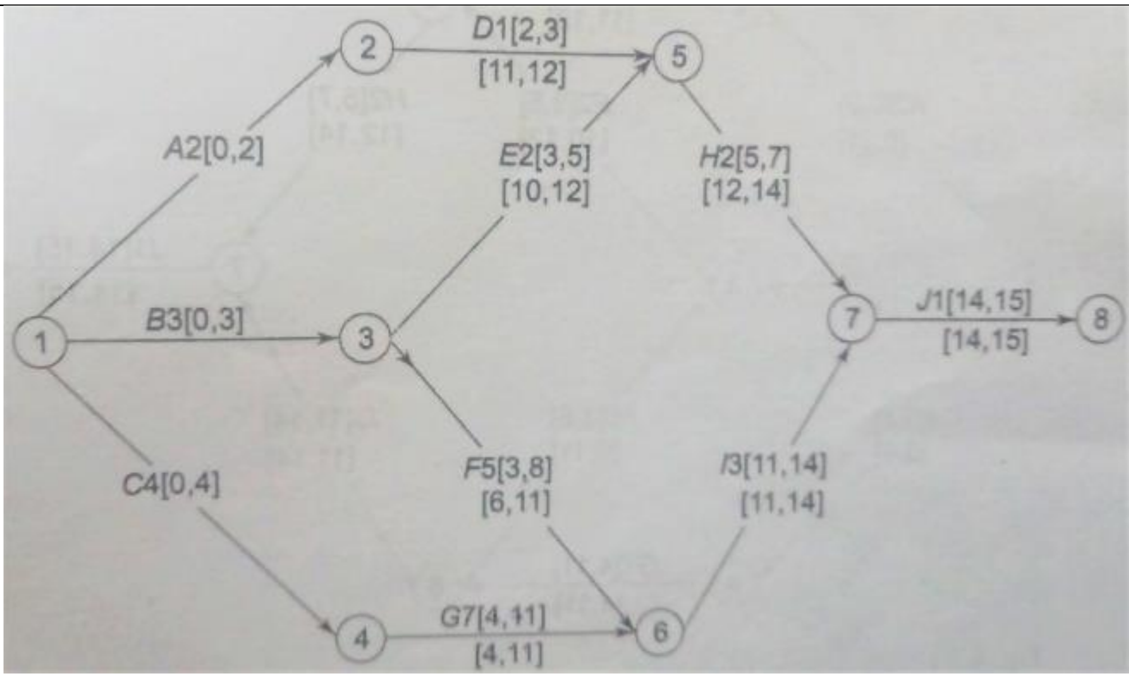


Figure 14

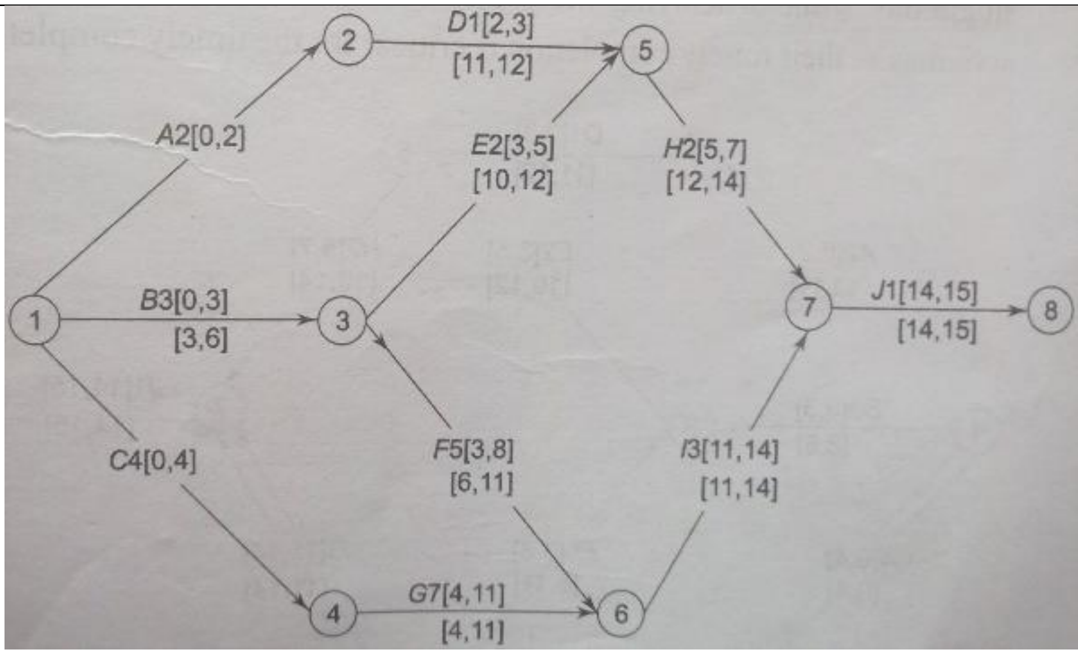


Figure 15

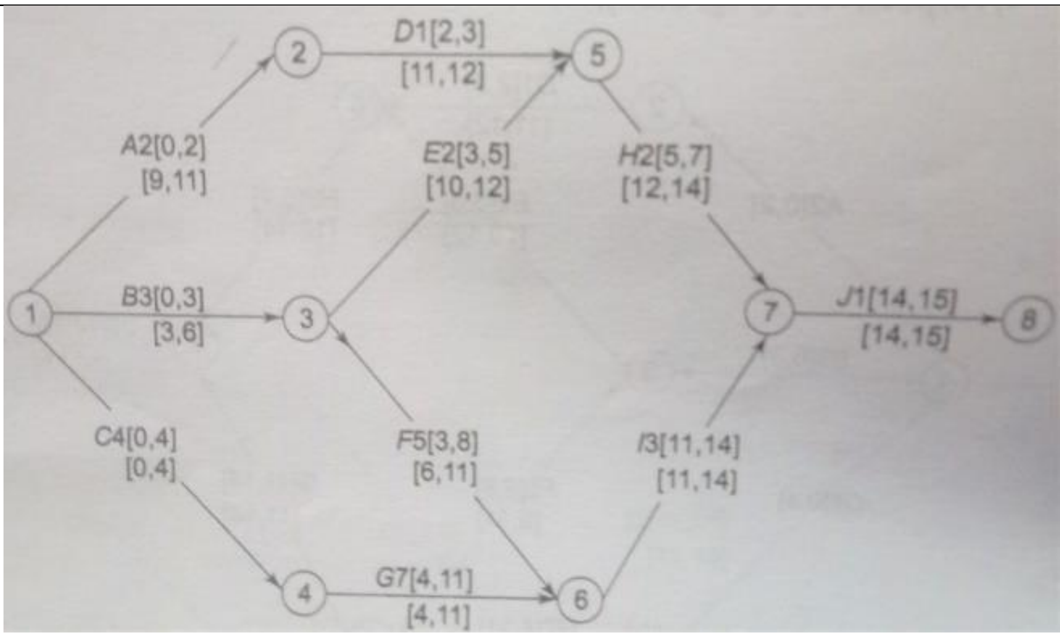


Figure 16

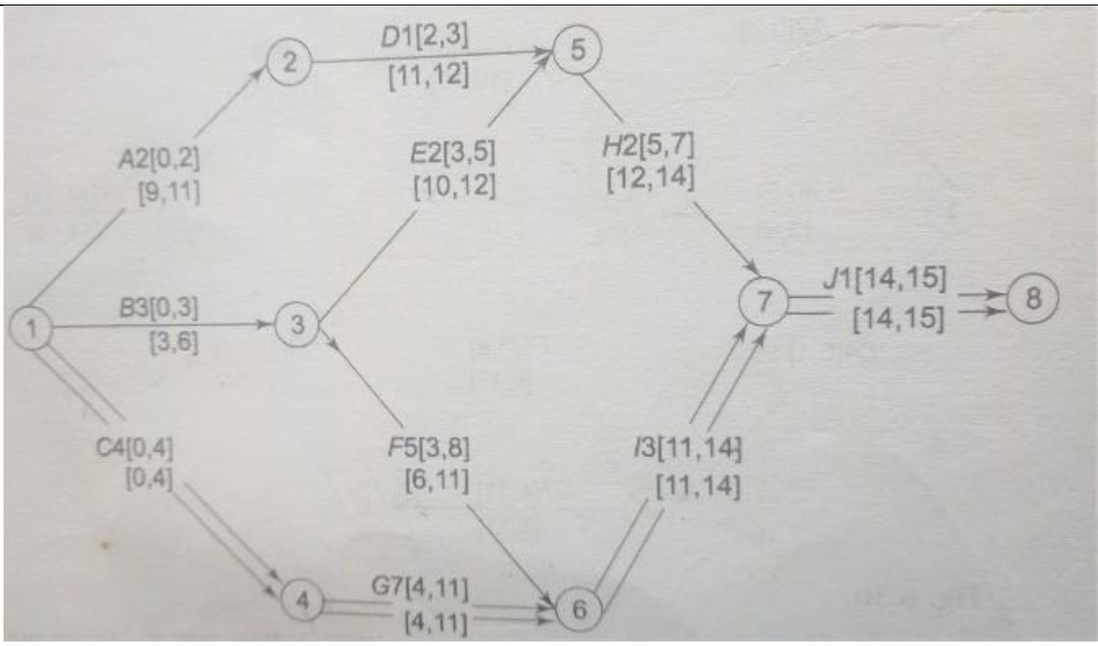


Figure 17

Limitation of CPM

CPM has operated successfully to identify the path with longest time period required. However, CPM has major limitation as it cannot be useful for the project having uncertainty about the duration of the activities. In above example, we were confident about the completion of each activity. Now just think of the situation if you are not sure about duration of activity A (Design House) with certainty, how would we calculate the path time and CPM ultimately. This limitation leads to development of the further method and it known as Programme Evaluation and Review Technique (PERT).

7.3 PROGRAMME EVALUATION AND REVIEW TECHNIQUE (PERT)

PERT is a method to analyze the involved tasks in completing a given project, especially the time needed to complete each task, and identifying the minimum time needed to complete the total project. PERT is applicable in the project when the duration of the activities cannot be predicted with the exact time of completion.

PERT is based on the assumption that an activity's duration follows a probability distribution instead of being a single value three time estimates are required to compute the parameters of an activity's duration distribution. Under PERT, it is assumed that the probability of completion of activity follows Normal Distribution (Z curve). With this understanding the time required to complete any activity can be divided into three major phases; it could be pessimistic time required, most likely time and optimistic time. The followings are the understanding about the same:

1. Pessimistic time (tp) – This is the longest possible time estimated for completion of activity. In other words, it is the time the activity would take if nothing did go well. It is based upon the premises that nothing will for right for the earliest completion of the activity. It means there will be all hindrances I the completion of the activity.

2. Most likely time (tm) – This is the time estimated for the completion of activity under normal circumstances. In other words, this is the time estimate that has highest probability of occurrence. This could be based on experience and expatriation the consensus best estimate of the activity's duration.

3. Optimistic time (to) - This is the shortest possible time estimated for completion of activity. In other words, it is the time the activity would take if things did go well. It is based upon the premises that everything will for right for the earliest completion of the activity.

These three estimates are combined together into single time estimates by taking a weighted average. This combined estimate is called the Expected time estimate (te) of the activity. A weight of 4 (weight is the importance attached to a value is attached to Most likely time (tm) and weights of 1 are attached to Pessimistic Time (tp) and Optimistic Time (to). Using the same, the calculation of Expected time (te) can be written as

$$te = (tp + 4tm + to) / 6$$

The rest procedure for identifying the critical path remains as it was with CPM method. With the objective to understand the same, let us consider the following example:

Activity	Node	to (in Days)	tm (in Days)	tp (in Days)
A	1-2	10	11	12
B	2-3	6	10	14
C	2-4	5	8	11
D	2-5	1	5	9
E	3-6	3	7	5
F	4-6	4	9	14
G	5-7	1	2	3
H	6-7	3	7	11
I	7-8	9	12	15
J	7-9	3	5	7

For activity A : te = (to + 4tm + tp)/6 = (10+44+12)/6 = 11

Similarly, calculate the expected time for all the activities and prepare a table with estimated time and the same is represented as under:

Activity	Node	te (in Days)
A	1-2	11
B	2-3	10
C	2-4	8
D	2-5	5
E	3-6	6
F	4-6	9
G	5-7	2
H	6-7	7
I	7-8	12
J	7-9	5

Based on this estimated time, the Critical Path can be evaluated using the above method of CPM and it is presented in following chart (refer figure 18)

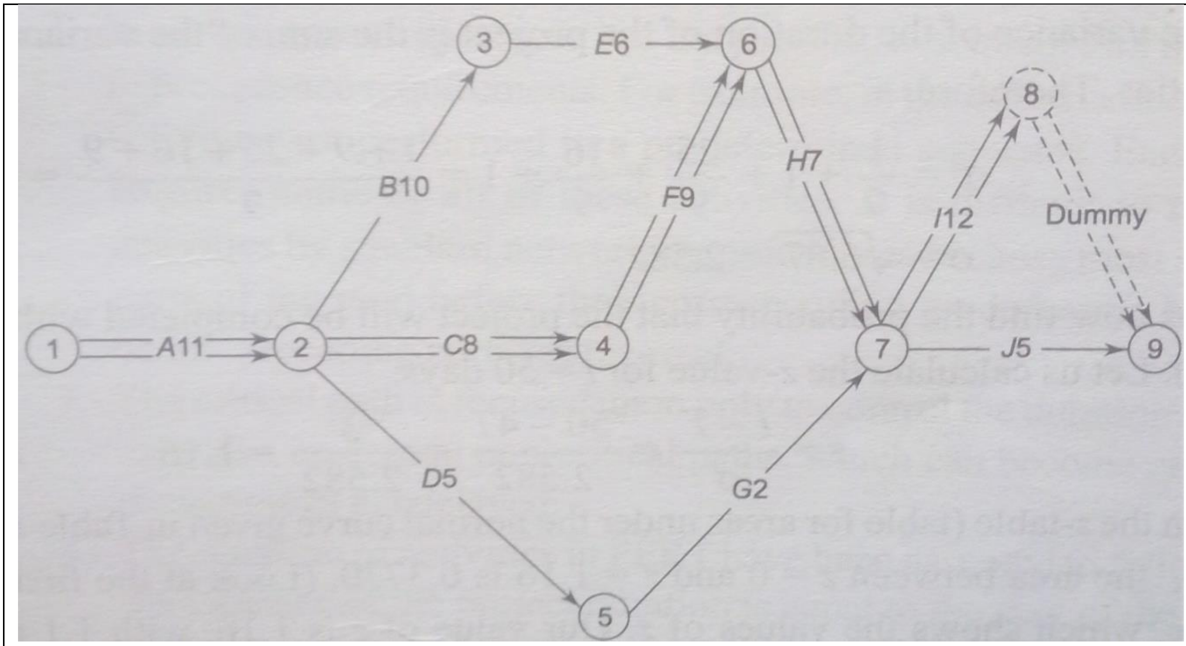


Figure 18

The expected duration for the given project is 47 days and the critical path is 1-2-4-6-7-8-9. Let us understand that the value of 47 days is expected duration and not exact duration. There is a possibility of having duration other than 47 and such variation is required to be calculated. Let us estimate the probability of completion of the project in 50 days(T), in stand of 47 days(te). There is four step procedure to estimate probability of completion of project within estimated time and the steps are as under:

Step 1 : Calculate expected duration for the project : We have already calculated expected duration and it is 47 days.

Step 2 : Calculate variance of the critical activities (Critical activity means an activity that is a part of Critical Path. In our case, it is activity A, C, F, H and I). Variance can be calculated as

$$\text{Variance } \sigma^2 = \{(to - tp)/6\}^2$$

Variance for Critical Activities are as Under:

For activity A : $\sigma^2 = \{(12\text{-}10) / 6\}^2 = (1/3)^2 = 0.1089$

For activity C : $\sigma^2 = \{(11\text{-}5) / 6\}^2 = (1)^2 = 1$

For activity F : $\sigma^2 = \{(14\text{-}4) / 6\}^2 = (5/3)^2 = 2.7889$

For activity H : $\sigma^2 = \{(11\text{-}3) / 6\}^2 = (4/3)^2 = 1.7629$

For activity I : $\sigma^2 = \{(15\text{-}9) / 6\}^2 = (1)^2 = 1$

The variance of the duration of the project is the sum of variances of all the critical activities. Therefore, For the project : σ^2 = 0.1089 + 1 + 2.7889 + 1.7629 + 1 = 6.67

So, σ = Square root of (6.67) = 2.582

Step 3 : Find the value of z (As per Normal Distribution),

Z = (te – T)/ σ = (50-47)/2.582 = 1.16

Step 4 : Using Z table, find out the probability of completion of the project.

To find out the probability of completion, one need to find the area between value of z =0 to z=1.16 from z table. We can get the value of z table using first two digit (here it is 1.1) as raw of table and the third digit (here it is 6) from the column and that brings the value of probability. Here it will be 0.3770. Add 0.5 to it to get the final value of probability of completion of the project in 50 days.

Therefore, the probability of completion of project in 50 days will be 0.3770 + 0.5 = 0.8770, i.e. 87.70% chances that the project will be completed in 50 days.

Differentiate between CPM and PERT

Point of difference	**CPM**	**PERT**
Orientation	It is activity oriented	It is event oriented
Nature	It is probabilistic	It is deterministic
Emphasis	CPM places dual emphasis on project time as well cost	PERT is primarily concerned with time only.
Usage	CPM is used for projects which are repetitive in nature and comparatively small in size	PERT is generally used for projects where time required to complete the activities is not known a priori.
No. of estimates	One time estimate is possible for activities (No allowance is made for uncertainty)	Three time estimates are possible for activities linking up two events

Benefits of CPM and PERT

1) Useful at many stages of project management

2) Mathematically simple

3) Give critical path and slack time

4) Provide project documentation

5) Useful in monitoring costs

Limitations of CPM and PERT

CPM and PERT have been criticised in the past for the following reasons:

1. It is really very difficult to estimate activities and their duration in real world, i.e. to clearly define the start and end points of the activities are not easy when it comes to practice. Even in case of complex projects, identifying the activities and their sequence is really not possible. In some cases, post planning stage, we may need to accommodate several activities and their duration for completion of the project.
2. The critical path focuses only on controlling the duration of the project. There may be existence of several non-critical paths, which may become critical upon the removal of certain activities or addition of certain activities. CPM and PERT completely ignores such possibilities.
3. In several complex projects, it is not possible to sequence all the activities according to precedence requirements. In the construction of Bridge, the various activities performed in predetermined sequence. Each phase of the bridge requires some or all of these activities. It is difficult to present them in standard network format.
4. The duration of the activities in PERT is assumed to be normally distributed, and the variance of the project duration is equal to the sum of the variances of all the critical activities.

7.4 MAXIMUM FLOW METHOD

In the daily life, there are numerous crucial problems that can be modeled as flow networks such as the flow of information through communication networks, current through electrical networks, liquids through pipes, calls in a communication network for particular services, parts through assembly lines, and it also emerges in process scheduling when allotting resources to processes in heterogeneous computing. Weighted directed connected graphs (a weighted directed connected graph is known as a network if the weight of each arc is a non-negative value) are usually utilized to represent these networks that model systems in which a material is created at the starting node (the one and only one that does not have entering edges and it is known also as the source), moves through the system, and at the target node (the one and only one that does not have leaving edges and it is also known as the sink) no longer exists.

Every edge has a capacity that is not negative and accepts the one whose quantity must not exceed its capacity. The flow of a flow network is a valued procedure defined on the links of the graph, i.e., assigning real numbers (non-negative) to the arcs fulfilling the following 2 conditions:

1. Capacity Constraint : The flow on every arc must not surpass its capacity and is non-negative. This is known as arc condition or capacity constraint
2. Flow Constraint : The incoming flow must equal the outgoing one at every node.

The maximum flow is the maximum incoming flow value to the sink node. According to the previous constraints, the overall flow to the sink node can be maximized. The following diagram depicts an instance of a flow network.

Getting the maximum flow of such a material in a flow network without breaking the restrictions of the capacity (nonnegative integer associated with every edge indicating the maximum allowed amount of such a material which can be sent through this edge and is also known as the weight of this edge) is known as the maximum flow problem. It is the problem of finding how much material can be transported between two locations through limited capacity ways, i.e., getting the uttermost overall flow from the source to the target in a directed weighted graph.

It is one of the classic combinatorial optimization problems, and numerous optimization problems can be reduced to it. It can be considered as a particular instance of network flow issues having increased difficulty like the circulation issue. It has been investigated by numerous researchers from experimental and theoretical perspectives utilizing various techniques since it models various important applications and is encountered in numerous crucial applications. Such as designing the transportation paths (deciding on the maximum number of travelers who can be moved between two stations per day, people in a public transportation system, traffic on roads, and other transportation networks as rail, and highway ones), shipping industry (organizations have to keep particular quantities of stocks in various warehouses and from the producer to the customers, distributing water to towns (deciding on how much fluids within pipes), traffic in computer and telephone networks, and its applications in other numerous fields such as logistics, computer science, bioinformatics, security, engineering, operations research, biological and logistic networks, and the allotment of hardware resources. In addition, it has numerous interesting special cases such as the bipartite graphs, which in turn model numerous natural flow problems.

- **Example for Maximum Flow Problem**

Let us consider the following network to analyse Maximum Flow Possible. Node 0 indicates the source (start point) of flow and node 5 indicates the sink (end point) of flow. The objective of the Maximum flow problem is to allow the maximum movement of product/units from source to sink. In order to understand the optimum flow, let us consider the following steps:

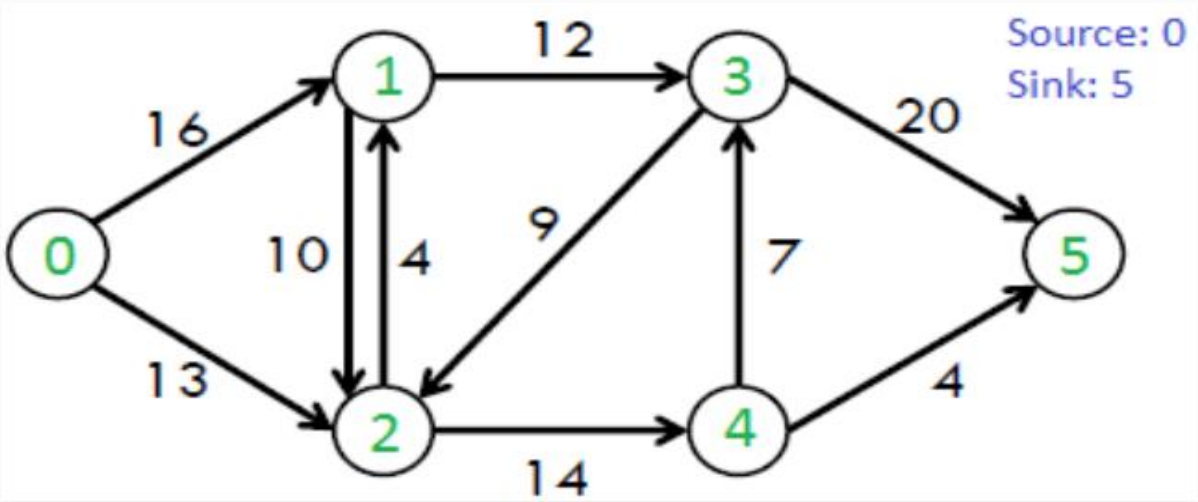
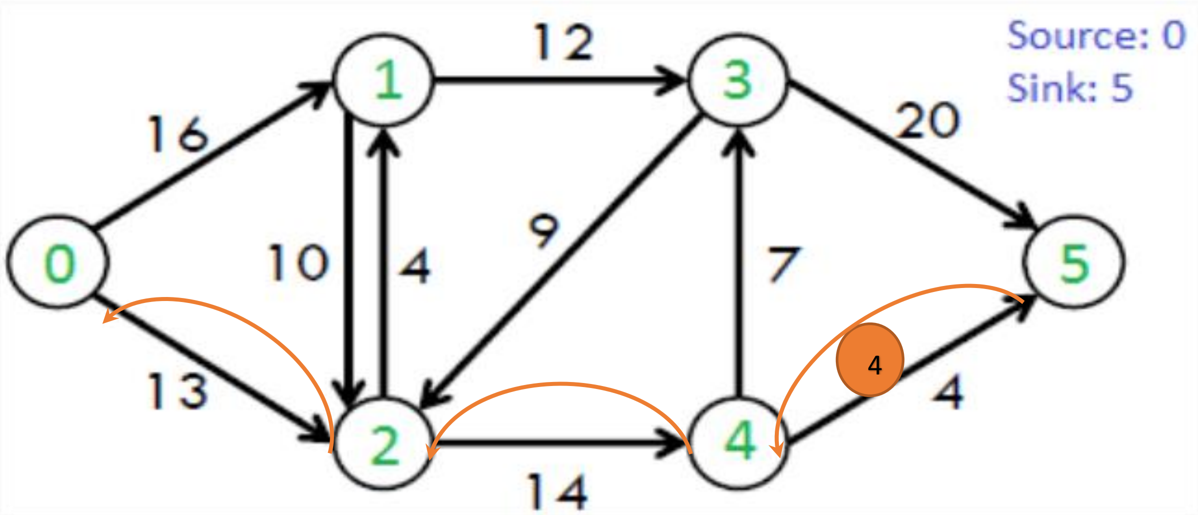


Figure 19

Step 1 : From Sink, identify the minimum amount of flow and try to shift the flow from sink to source. Here in our example flow of 4 unit is the minimum flow from sink (Node 5) that can be shifted initially under this step.



Step 2 : From Sink, identify the second minimum from and try to shift the same to source. Here in our example, flow of 20 units is the minimum flow from sink (Node 5) that can be shifted under this step (refer figure 20)

Flow of 20 units can be shifted from node 5 to node 3. At node 3, the flow can be divided into two parts, node 3 to 1 and node 3 to 4. According to flow constraint, node 3 to 1 can flow only 12 units. Therefore, let us try to accommodate remaining units with node 3 to 4 that can flow additional 7 units. It means the maximum capacity of node 3 to flow units is 19 (12+7) units only. And hence the flow of 20 units cannot be continued. Therefore, the maximum flow of 19 units can flow across from sink to source under step 2 and the revised flow diagram is given in figure 21.

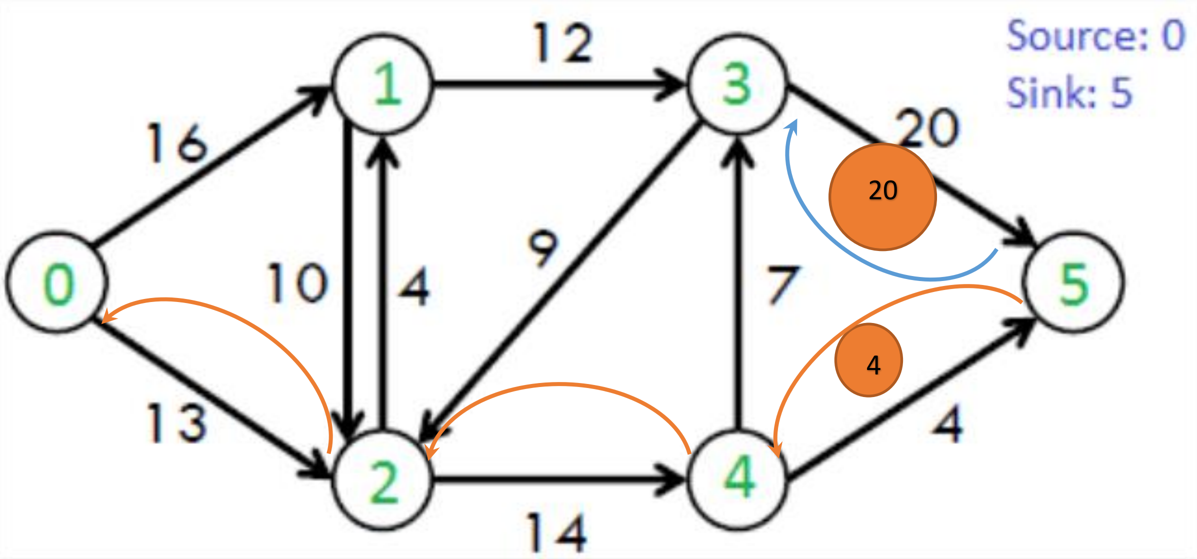


Figure 20

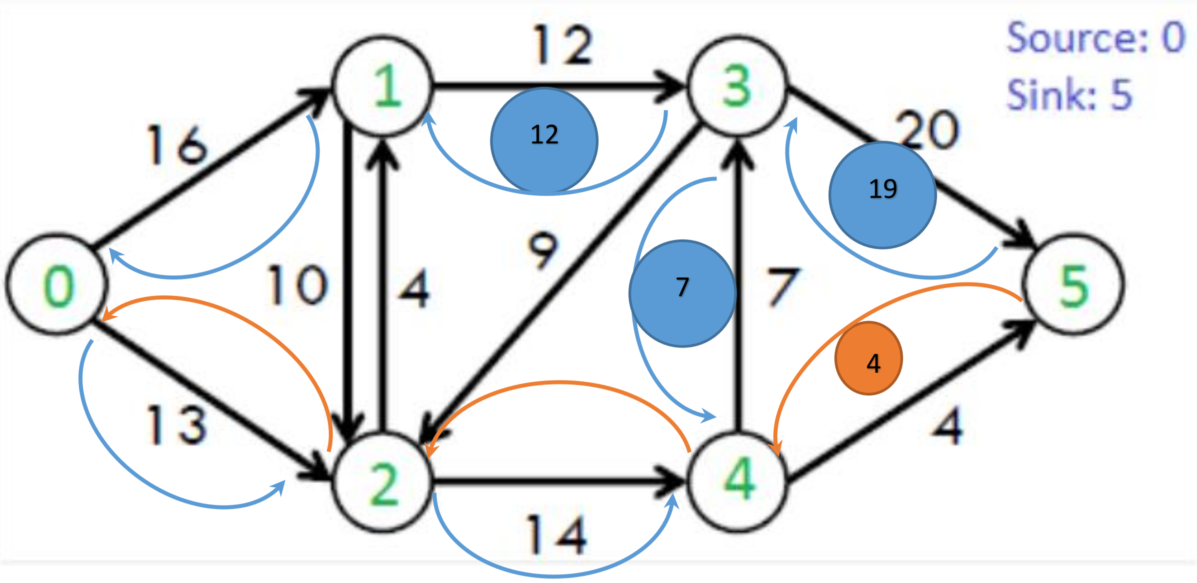


Figure 21

Step 3 : Identify the possibility of additional flow, if any within the network. In our case, the network is operated at maximum capacity (refer figure 22). Therefore, for given network the maximum possible flow is 23 units from source to sink.

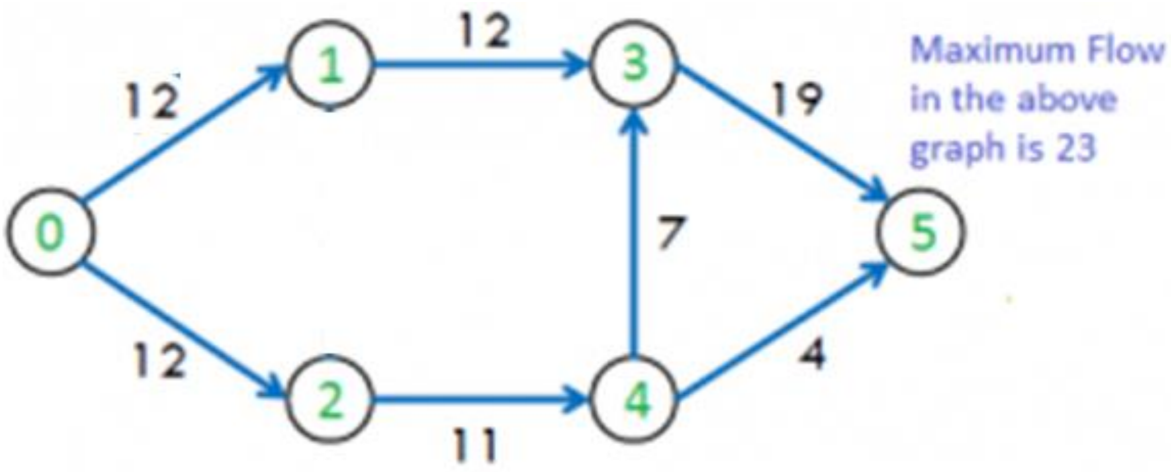


Figure 22

Application of Maximum Flow Method

Many vital applications in a variety of domains can be handled as a maximum flow problem. That is why it has been handled with the use of a variety of approaches. This was researched in this work; the Ford Fulkerson, the push relabel, and many recent approaches. The Ford Fulkerson can be applied easily for tackling a variety of real-life applications, as it is a simple method having straightforward steps. In fact, it is considered a game changer in the graph approaches.

This is obvious from its recent crucial real-life applications aforementioned. Also, the push relabel has its recent applications as well. The difference between the push relabel and augmenting algorithms is rather small. As nature inspired metaheuristics have been utilized successfully in dealing with a variety of crucial applications, they have been utilized for handling the maximum flow problem in various domains as explained previously. Further, hybrid and parallel approaches are increasingly used for tackling a variety of optimization problems and hence, they are used for handling the maximum flow problem.

❖ CHECK YOUR PROGRESS

- **Answer the following questions in detail.**

1. Compare CPM with PERT.
2. Explain applications of Maximum flow method.
3. How to calculate the probability of completion of project within given time frame under PERT? Explain with example.
4. Describe with example : Critical Path Method
5. Discuss various Network analysis methods.

- **Answer the following Multiple Choice Questions**

1. CPM Stands for ________________

   A. Crucial Path Measurement B. Critical Path Measurement

   C. Critical Path Method D. Crucial Path Method

2. Task performance in CPM is popularly known as ________________

   A. Activity B. Event C. Dummy D. Role

3. The starting point for Maximum flow is popularly called ___________

   A. Source B. Sink C. STP D. Upper Point

4. What is PERT analysis based on?

   A. Optimistic measures B. Most likely measures

   B. C.Pessimistic measures D. All of the above

5. Activity in a network diagram is represented by?

   A. Circle B. Square C. Arrow D. Rectangle

Answer

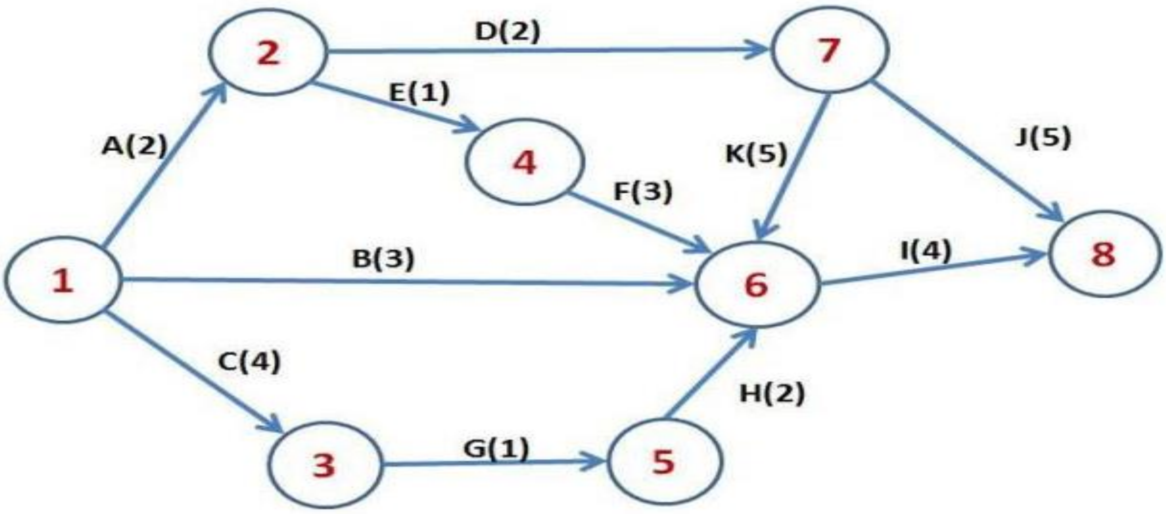
(1) C (2) A (3) A (4) D (5) C

Difference

1. Differentiate between CPM and PERT

Sums (Practical)

1. Circus are very popular among the children in a town. Gulbarga is a small town located in Chhatisgarh state of India. A circus playing company decides to establish canopy for Circus and stay in the town for around 4 Months. The following is a network diagram of activities for Circus canopy with time period (In No of Days). Identify the critical path and calculate the most no. of days required for canopy.



2. The following details are available regarding a construction of Dam:

Activity	Predecessor Activity	Duration (Weeks)
A	-	3
B	A	5
C	A	7
D	B	10
E	C	5
F	D,E	4

Determine the critical path, the critical activities and the project completion time.

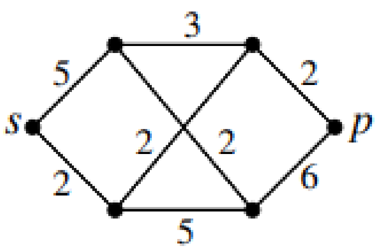
3. Draw the network diagram and determine the critical path for the following project:

Activity	Time estimate (Weeks)
1- 2	5
1- 3	6
1- 4	3
2 -5	5
3 -6	7
3 -7	10
4 -7	4
5 -8	2
6 -8	5
7 -9	6
8 -9	4

4. For the given activities determine:
   1. Critical path using PERT.
   2. Calculate variance and standard deviation for each activity.
   3. Calculate the probability of completing the project in 26 days.

Activity	t_o	t_m	t_p
1-2	6	9	12
1-3	3	4	11
2-4	2	5	14
3-4	4	6	8
3-5	1	1.5	5

5. Identify the Maximum flow for the following network



UNIT – 8 WAITING MODEL (QUEUING THEORY)

8.1 INTRODUCTION

Queuing theory is a branch of mathematics that studies and models the act of waiting in lines. Queueing models provide the analyst with a powerful tool for designing and evaluating the performance of queueing systems. The Queuing Theory is concerned with studying all the various dynamics of lines – or "queues" – and how they may be made to operate more efficiently. It is essentially the study of "waiting in line," including how people behave when they have to queue up to make a purchase or receive a service, what types of queue organization move people through a line most efficiently, and how many people can a specific queuing arrangement process through the line within a given time frame. Queuing theory is also applied to moving objects or information through a "line". For example, an auto manufacturer may look to queuing theory for guidance on the most efficient way to set up its assembly lines. A transport company, such as Delhivery or BlueDart, can use queuing theory to determine the most operationally efficient manner of transferring packages from one transport vehicle to another.

In the world of business, queuing theory can help a company's executives determine the best way to set up and organize Business Operations so as to maximize both sales and customer service satisfaction.

Financial analysts may construct models based on queuing theory to make projections about how changing an operational variable may improve queuing efficiency and, as a result, bottom-line profitability. The use of queuing theory has become so popular that there are now there are several online queuing calculators help us in doing a basic analysis of a given queuing setup.

8.2 BASIC CONCEPTS OF QUEUING

Queuing theory is essentially a vehicle for cost analysis. It would be prohibitively expensive, or indicative of not having very many customers, for most businesses to operate in a manner so that none of their customers or clients ever had to wait in line. As a simplistic example, for a movie theatre to eliminate the circumstance of people having to wait in line to purchase a movie ticket, it would likely need to set up fifty to a hundred ticket booths. However, the theatre obviously could not afford to pay a hundred ticket sellers.

Therefore, businesses use information gleaned from queuing theory in order to set up their operational functions so as to strike a balance between the cost of servicing customers and the inconvenience to customers caused by having to wait in line. The basics of queuing include the people waiting in line and the performance of the service that they're waiting to receive. In studies on queuing, it is usually broken down into four categories, as follows.

1. Arrival – The process by which customers arrive at the line/queue
2. The Queue – The nature or operation of the queue itself (How does the line move along?)
3. Service – The process of providing the service that a customer is waiting on (for example, being seated and then being served in a restaurant – note that the restaurant has to consider the dynamics of two separate queues, the queue of people waiting to be seated, and then the queue of people already seated who are waiting to be served. The latter might be further broken down into the two queues of waiting to have your order taken and waiting for your food to arrive at your table)
4. Leaving – The process of departing from the queue location (for example, businesses that offer a drive-through service have to consider how people leaving the drive-through may impact people entering the business' parking lot)

Now let us understand above four factors of queuing in details:

1. Arrival – Factor 1 of Queuing model

Queuing models analyze the operational aspects and variables involved in each of the four categories of queuing outlined above. Following are some of the variables that can affect the functioning and operational efficiency of each part of a queue, and that, therefore, should be considered by the business where a queue forms. Factors to consider in relation to the arrival of people at the queuing location include such things as the number of people, on average, who arrive within a given time frame, such as one hour. A related factor is that of substantial fluctuations in the amount of traffic/arrivals that occurs at different times of the day and/or on different days of the week or month.

For Example: Grocery store knows that in order to avoid queues getting backed up, they need to have more employees working during rush hour on a Friday than, say, on Wednesday mornings between 10 a.m. and noon.

2. The Queue – Factor 2 of Queuing model

How does the line move along? For example, does it work better for a bank to have just one line of customers waiting for the next available teller or cashier, or to have separate lines for each teller? Characteristics of human behavior become an important part of queuing theory when posing such a question. While one line of customers being fed to four different teller stations versus four separate lines at each teller station may not have a significant effect on how quickly or efficiently customers are served, it may well have an impact on customer satisfaction.

Although ultimately, the wait time to be served may be roughly the same regardless of the line arrangement, customers may *feel*, or perceive, that they are being served more quickly if they only have to wait in line behind two or three people (each teller station has its own queue) as opposed to having to stand in line behind 10 or 12 people (one line of customers being fed to all four teller station). There are also basic practicalities to consider: If the business office is relatively small, will using just a single line result in a line so long that it extends back out the door? Many people seeing a situation like that may well be discouraged from doing business there. They may instead choose to go to a competitor that appears to offer less wait time.

Note the part about "appears" to offer less wait time. Doing business with the competitor may, in fact, involve approximately the same amount of time waiting in line. The only difference may be that the competitor chose to go with separate lines for each service station rather than one single line for all the stations, thus avoiding having a line that extends back out the door. Here, you can see that there are aesthetics of queues to be considered in addition to any operational efficiency factors.

3. Service – Factor 3 of Queuing model

here are also variables that exist in relation to the actual provision of service. A common example is the "express lane" in grocery stores, reserved for customers who are only purchasing a small number of items. (Typically, express lanes are designated for customers with "12 items or less" or "20 items or less"). The reason such express lanes exist is that grocery stores using queuing theory have found that customer satisfaction is improved by enabling customers who are only buying a few things to check out more quickly, as opposed to having to wait in line behind other customers with full carts of groceries.

Other factors that impact actually providing service include how long it takes to provide service to each customer or client, the number of servers required for maximum operational and cost efficiency, and the rules governing the order in which customers are served. While most queues operate on a "first-come, first-served" basis, it is not appropriate for some businesses.

For Example: the waiting area at a hospital emergency room. Rather than using a "first arrival" basis for service orders, patients are served based on the severity of their illness or injury. It necessitates adding a service step known as "triage", whereby a nurse evaluates each patient in terms of the severity of their emergency to decide where in the line of receiving service that patient is placed.

4. **Leaving – Factor 4 of Queuing model**

The elements associated with customers departing a queue location are commonly basic logistical matters. The example was related above of how businesses with drive-through operations have to take into account how people leaving the drive-through may affect incoming traffic to the location.

For example: A restaurant is determining whether to have servers present bills and collect payment at a customer's table or to have customers pay their bill to a cashier on their way out.

8.3 CHARACTERISTICS OF QUEUEING SYSTEMS

Any queuing system possesses several characteristics and are presented as follows:

1. Calling population: the population of potential customers, may be assumed to be finite or infinite and that leads to two types of Queue.
2. Finite/Infinite Queue: The queuing system operates on this basic fundamental of finite of infinite queue. For Example: Queue at Railway station ticket window or at Hospital emergency ward are typical examples of infinite queue. While queue at Bank or Restaurant are example of finite queue. Infinite queue requires server to operate for infinite time.
3. No. of Queues: This is another characteristics of Queue. The system may be single queue system or multiple queue system. The multiple queue system can also be

further classified as Parallel queue or merging queue. The further section on types of queue details the same.

4. Customers: Customer refers to anything that arrives at a facility and requires service, e.g., people, machines, trucks, emails, packets, frames.
5. Server: refers to any resource that provides the requested service, e.g., repairpersons, machines, runways at airport, host, switch, router, disk drive, algorithm
6. No. of Servers: The system works with single server of multiple servers.
7. System Capacity: a limit on the number of customers that may be in the waiting line or system. Based on system capacity, the system are characterised as Limited Capacity and Unlimited capacity;

- Limited capacity, e.g., an automatic car wash only has room for 10 cars to wait in line to enter the mechanism. In this case, if system is full no customers are accepted anymore.
- Unlimited capacity, e.g., concert ticket sales with no limit on the number of people allowed to wait to purchase tickets.

8. Arrival Process: It reflects the pattern of arrival of the customers. For finite queuing system, arrival is not a major point of concern as the system knows that the arrival is finite. But for infinite there are two possible arrival types:

- Random arrivals: interarrival times usually characterized by a probability distribution. The Poisson arrival process (with rate λ), where a time represents the interarrival time between customer n-1 and customer n, and is exponentially distributed (with mean $1/\lambda$).
- Scheduled arrivals: interarrival times can be constant or constant plus or minus a small random amount to represent early or late arrivals. For Example: Social worker has scheduled his meeting with Chief Minister Office or scheduled airline flight arrivals to an airport.

9. Queue behaviour: the actions of customers while in a queue waiting for service to begin, for example:

- Balk: leave when they see that the line is too long
- Renege: leave after being in the line when its moving too slowly
- Jockey: move from one line to a shorter line

10. Queue discipline: the logical ordering of customers in a queue that determines which customer is chosen for service when a server becomes free, for example:

- First-in-first-out (FIFO)

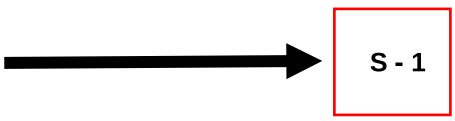
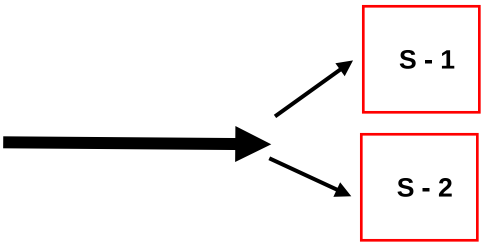
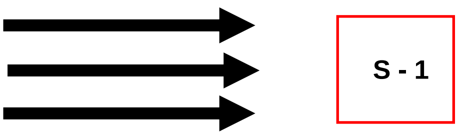
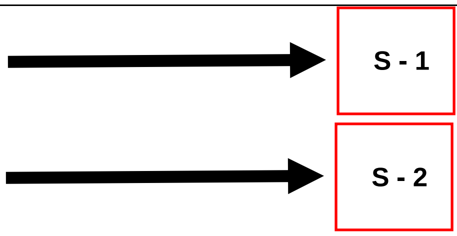
- Last-in-first-out (LIFO)
- Service in random order (SIRO)
- Shortest processing time first (SPT)
- Service according to priority (PR)

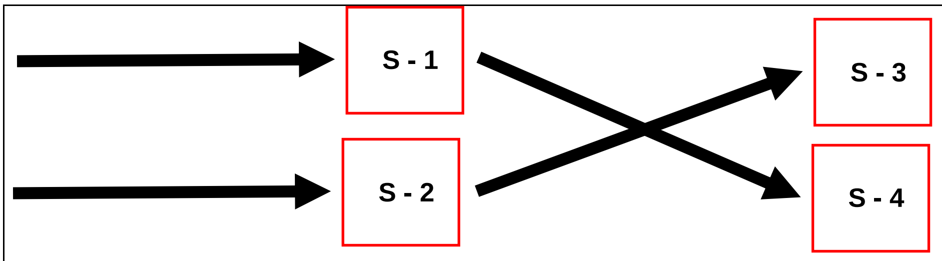
8.4 TYPES OF QUEUING SYSTEM

There are several Queuing systems based on no of queue and servers:

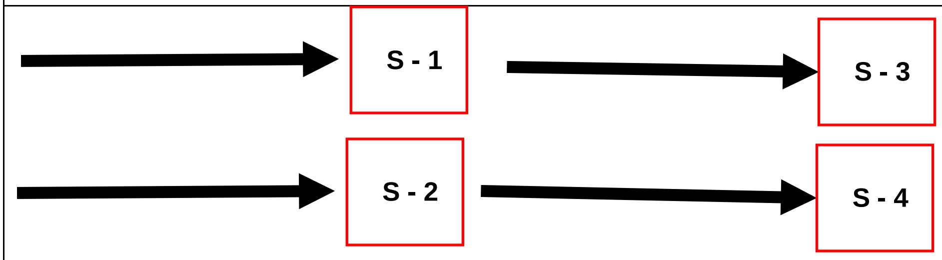
1. Single Queue, Single Server
2. Single Queue, Multiple Servers
3. Multiple Queue, Single Servers
4. Multiple Queue, Multiple Servers
5. Multiple Queue, Multiple Servers with Zigzag Pattern
6. Multiple Queue, Multiple Servers with Straight Pattern

The following section picturised the above types of queue with possible real life examples.

 **Single Queue, Single Server** Ex. : Fees Payment Queue at College/University	 **Single Queue, Multiple Server** Ex. : Airport for Check in of the luggage
 **Multiple Queue, Single Server** Ex. : Queue for Prayer and Darshan at Sai Baba/any other Temple	 **Multiple Queue, Multiple Server** Ex. : General Ticket window at Railway station



Multiple Queue, Multiple Server with Straight Pattern
Ex. : Admission Counters and Fees Payment for Offline Admission Process



Multiple Queue, Multiple Server with Straight Pattern
Ex. : Process for Document Verification and other formality at Passport Office

8.5 SINGLE SERVER MODEL (M/M/1/A)

The above symbol can be read as Server capacity to work at aa time is 1 unit and no. of customers called is α units. Before the starting of the understanding, let us understand several important terminologies:

Arrival Rate: The rate at which customers arrives in system. The rate is to be measured with respect to time. It can be measured as per Second, per Minute, per Hour, per Day, per Week, per Month or per Year. It is denoted by Lamda (λ).

Service Rate: The rate at which service performs the services. The rate is to be measured with respect to time, as mentioned above in case of arrival rate. It is denoted by Mu (μ).
Waiting time: It is the time required for the customer in queue before he/she/it is served.
Serving Time: It is total time spent by customer in queuing system. It includes time in waiting queue and time to serve him/her/it.

Server Utilization Rate: the proportion of time that a server is busy. There are really very few systems where server is busy for 100 percentage of time. Even in real world when server is human being several human factors such as thrust, hunger and fatigue play a role and reduce server utility. Therefore, the server utilisation rate is to be calculated and the effort is kept to enhance it. It is denoted by Rho (ρ).

To begin with numerical, we will use the following notations:

λ: mean rate of arrival

μ: mean service rate

ρ: Service utilisation rate

W: mean waiting time in the system

Wq: mean waiting time in the queue

L: mean number of customers in the system

Lq: mean number of customers in the queue

Pn: probability that there are n customers in the system

A few important formulas are as under:

$P = \lambda / \mu$

Let us have some of the numerical:

Illustration 1.: Customers arrive at sales counter managed by single person according to poisson distribution with average arrival rate of 20 customers per hour. The time required to serve customer is average 100 seconds. Find the average waiting time of a customer in queue and in system. Also find the service rate. In one hour for how many minutes, the sales person at counter is free/unoccupied.

Solution: Before we start with calculation, convert both arrival rate and service rate in no. of customers per hour.

Arrival rate (λ) = No. of customers arrived per hour = 20

Service rate (μ) = No. of customers served per hour = No of seconds in an hour/time per customer = 3600/100 = 36

(1) Average waiting time in Queue = Wq = λ / μ (μ - λ)

= 20 / 36 (36-20)

= 0.03472 Hours

(2) Average waiting time in System = Ws = 1 / (μ - λ)

= 1 / (36-20)

= 1 / 16 = 0.0625 Hours

(3) Service Rate = ρ = λ / μ = 20 / 36 = 55.56%

(4) No. of minutes sales person is free = 1 – ρ = 1 – 0.5556

= 0.4444 Hours = 0.4444 * 60

= 26.66 Minutes

Illustration 2.: A departmental store has a single cashier. During the rush hours, customers arrive at the rate of 20 customers per hour. The average customers that the cashier can serve in an hour is 24 per hour. Find out,

(1) Probability that the cashier is idle
(2) Probability that a cashier is busy
(3) Average No. of Customers in queue
(4) Average No. of customers in the system
(5) Average time a customer spent in the system
(6) Average time a customer spent in the queue

Solution: Before we start with calculation, convert both arrival rate and service rate in no. of customers per hour.

Arrival rate (λ) = No. of customers arrived per hour = 20

Service rate (μ) = No. of customers served per hour = 24

Service rate (ρ) = λ /μ = 20/24 = 0.8333

(1) Probability that the cashier is idle = 1 – ρ = 1 – 0.8333 = 0.167 = 16.7%

(2) Probability of cashier is busy = ρ = 0.8333 = 83.33%

(3) Average No. of customers in the queue = $L_q = \lambda^2 / \mu (\mu - \lambda)$

$= (20)^2 / 24 (24-20)$

= 400 / 24*4 =4.166

= 4 customers (Approx.)

(4) Average No. of customers in the system = $L_s = \lambda / (\mu - \lambda)$

= (20) / (24-20)

= 20 / 4 = 5 Customers

(5) Average waiting time in the system = $W_s = 1 / (\mu - \lambda)$

= 1 / (24-20)

$= 1 / 4 = 0.25$ Hours

(6) Average waiting time in the queue = Wq = $\lambda / \mu (\mu - \lambda)$

$= 20 / 24 (24-20)$

$= 20 / 24 * 4 = 0.2083$ Hours

❖ CHECK YOUR PROGRESS

- **Answer the following questions in detail.**

1. Explain “Types of Queue”
2. Explain with example, application of queuing system
3. What is the meaning of Service rate and Probability of waiting in case of queuing?
4. Describe any two characteristics of Queuing system
5. Explain in detail; Arrival and Service in case of Queuing

- **Answer the following Multiple Choice Questions.**

1. **Arrival in queue represents _______________**

   A. Addition of new customer B. change in queue length C. Increase in length of system D. All of the above

2. **Those customers who move from one line to a shorter line are known as _______________**

   A. Renege B. Balk C. Jockey D. Role

3. **Those customers who leave after being in the line when its moving too slowly are popularly called __________**

   A. Renege B. Balk C. Jockey D. Role

4. **Those customers who leave when they see that the line is too long are known as _____________**

   A. Renege B. Balk C. Jockey D. Role

5. **FIFO stands for _____________________**

   A. First In First Out B. First In First Overflow C. First in Fast Out D. Full in Full Out

Answer

(1) D (2) C (3) A (4) B (5) A

- **Difference**

1. Differentiate between Single Server and Multiple Server
2. Differentiate between Single Queue and Multiple Queue

- **Sums (Practical)**

1. In a railway yard, 10 trains arrive a day for getting unloaded, on an average. The unloading team can work for whole day and can unload 16 trains a day. You are required to calculate (1) Average no. of trains in system for unloading (2) Percentage of time the unlading team is unoccupied (3) Average waiting time for train in queue
2. In a restaurant, there 30 customers arrived in an hour. Assuming only one servicing counter that takes 1.5 minutes to serve a customer on an average. Calculate (1) Probability of having no customer in waiting (2) Average length of queue
3. A car washer in the City mall serves car for washing. On any given day, there are 150 cars arrived for washing. A car washing workshop has one washing station and can wash 200 cars a day. If a customer arrived to wash his car, what is the probability that the car washing station is empty.

- **List of Formulae**

(1) Probability of Queue being busy = $\rho = \lambda / \mu$

(2) Probability of empty Queue = $1 - \rho$

(3) Average No. of customers in the queue = $Lq = \lambda^2 / \mu (\mu - \lambda)$

(4) Average No. of customers in the system = $Ls = \lambda / (\mu - \lambda)$

(5) Average waiting time in Queue = $Wq = \lambda / \mu (\mu - \lambda)$

(6) Average waiting time in the system = $Ws = 1 / (\mu - \lambda)$

UNIT 9 THEORETICAL PROBABILITY DISTRIBUTION

9.1 INTRODUCTION

In the real world, we rarely come across experiments with single outcomes like heads or tails. Generally, we do the experiment as a set of events and carry it for a number of times which give us a collection of outcomes which we can represent in the form of theoretical distribution. By theoretical distribution, we take mean of a frequency distribution, which we obtain in relation to a random variable by some mathematical model.

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 extensive categories, viz. discrete probability distribution and continuous probability distributions, depending upon whether the random variable is discrete or continuous. A random exponent is assumed as a model for theoretical distribution, and the probabilities are given by a function of the random variable is called **probability function**. For example, if we toss a fair coin, the probability of getting a head is ½. If we toss it for 20 times, the probable number of getting a head is 10. We call this as the theoretical or expected frequency of the heads. But actually, by tossing a coin, we may get 10, 15 or 18 heads which we call as the observed frequency. Thus, the observed frequency and the expected frequency may equal or may differ from each other due to fluctuation in the experiment. There are a large number of distributions in each category; we shall discuss only some of them having important business and economic applications.

9.2 IMPORTANCE

With theoretical probability, you don't actually conduct an experiment (i.e. roll a die or conduct asurvey). Instead, you **use your knowledge about a situation, some logical reasoning, and/or known formula to calculate the probability of an event happening**.

A distribution that is derived from certain principles or assumptions by logical and mathematical reasoning, as opposed to one derived from real-world data obtained by empirical research. Examples include the normal distribution, binomial distribution, Poisson distribution, Hyper Geometric distribution, Uniform distribution, etc.

There are several reasons why the field of Theoretical Probability exists. Sometimes, conducting an experiment isn't possible for practical or financial reasons. For example, you might be studying a rare genetic trait in salamanders and you want to know what the probability of any one salamander having the rare trait is. If you don't have access to all of the salamanders on the planet, you won't be able to conduct an experiment so you'll have to rely on theory to give you the answer. Theoretical probability is also used in many areas of science where direct experimentation isn't possible. For example, probabilities involving subatomic particles or abstract structures like vector spaces.

Probability provides information about the likelihood that something will happen. Meteorologists, for instance, use weather patterns to predict the probability of rain. In epidemiology, probability theory is used to understand the relationship between exposures and the risk of health effects.

Discrete Variable: A discrete variable is a variable whose value is obtained by counting. For example number of students present, number of red marbles in a jar, number of heads when flipping three coins, students' grade level, etc.

Continuous Variable: A continuous variable is a variable whose value is obtained by measuring. For example height of students in class, weight of students in class, time it takes to get to school, distance traveled between classes, etc.

Random Variable: A random variable is a variable whose value is a numerical outcome of a random phenomenon.

- A random variable is denoted with a capital letter
- The probability distribution of a random variable X tells what the possible values of X are and how probabilities are assigned to those values
- A random variable can be discrete or continuous

9.3 DISCRETE PROBABILITY DISTRIBUTION

There are basically two types of random variables, called continuous and discrete random variables. We shall discuss the probability distribution of the discrete random variable. The discrete random variable is defined as the random variable that is countable in nature, like the number of heads, number of books, etc.

A discrete random variable X has a countable number of possible values. Such kind of variable has probability distribution which is known as probability mass function or discrete probability distribution. Some important discrete probability distributions are listed below. We learn about it.

The probability distribution that deals with discrete random variable is called the probability mass function (pmf). There are various types of discrete probability distribution. They are as follows:

A random variable X is said to have a discrete probability distribution called the discrete uniform distribution if and only if its probability mass function (pmf) is given by the following:

$$(x) = \begin{cases} \frac{1}{\div}\, x = 1,2,3,\ldots,n \\ 0; \quad elseshere \end{cases}$$

A random variable X is said to have a discrete probability distribution called the Bernoulli distribution if and only if its probability mass function (pmf) is given by the following:

$$(x) = \begin{cases} (1-p)^{1-x};\ x = 0, 10; \\ elseshere \end{cases}$$

A random variable X is said to have a discrete probability distribution called the Binomial distribution if and only if its probability mass function (pmf) is given by the following:

$$P(x) = \begin{cases} nC_x p^x q^{n-x}; & x = 0,1,2,\ldots,n, 0 < p < 1, q = 1-p \\ 0; & elsewhere \end{cases}$$

A random variable X is said to have a discrete probability distribution called Poisson distribution if and only if its probability mass function (pmf) is given by the following

$$P(x) = \begin{cases} \dfrac{e^{-m}m^x}{x!}; & x = 0,1,2,\dots,n \\ 0; & elsewhere \end{cases}$$

A random variable X is said to have a discrete probability distribution called the geometric distribution if and only if it is the following:

$$P(x) = \begin{cases} q^x p; & x = 0,1,2,\dots,n, 0 < p < 1, q = 1 - p \\ 0; & elsewhere \end{cases}$$

A random variable X is said to have a discrete probability distribution called the hyper geometric distribution with the parameters N, M and n if it assumes only non-negative values with the probability mass function as the following:

$$P(x) = \begin{cases} \dfrac{MC_x\ (N-M)_{-k}}{NC_n}; & k = 0,1,2,\dots,\min(n, M) \\ 0; & elsewhere \end{cases}$$

We note many discrete probability distributions with their probability mass function (pmf). In this chapter we learn more about Binomial Distribution, Poisson Distribution and Hyper Geometric Distribution in detail.

9.3.1 Binomial Probability Distribution:

Binomial distribution was first given by a **Swiss** mathematician **James Bernoulli** in the beginning of eighteenth century. When an experiment can result only in two ways success and failure, and when such an experiment can be repeated independently and in each trial probability of success remains the same, then such trials are known as Bernoulli trials. Tossing of a coin for n times is an example of Bernoulli trials.

The binomial distribution is a probability distribution that summarizes the likelihood that a value will take one of two independent values under a given set of parameters or assumptions. The underlying assumptions of the binomial distribution are that there is only one outcome for each trial, that each trial has the same probability of success, and that each trial is mutually exclusive, or independent of one another. The binomial distribution is a common discrete distribution used in statistics, as opposed to a continuous distribution, such as the normal distribution. This is because the binomial distribution only counts two states, typically represented as 1 (for a success) or 0 (for a failure) given a number of trials in the data. The binomial distribution thus represents the probability for x successes in n trials, given a success probability p for each trial.

Binomial distribution summarizes the number of trials, or observations when each trial has the same probability of attaining one particular value. The binomial distribution determines the probability of observing a specified number of successful outcomes in a specified number of trials. The binomial distribution is often used in social science statistics as a building block for models for dichotomous outcome variables, like whether a Republican or Democrat will win an upcoming election or whether an individual will die within a specified period of time, etc.

Suppose an experiment is performed which can result in two ways success and failure. Suppose the probability of success is p and that of failure is q (p + q = 1). Suppose this experiment is repeated independently for n times and the probability of success in each trial remains constant i.e., p. If we write S for success and F for failure, then we have probability of success = P(S) = p and probability of failure = P(F) = q.

We shall find out the probability of getting x successes out of n such independent trials. x successes and $(n - x)$ failures can occur in the following way:

x times $\longrightarrow$ (n - x) times

S.S.S.......S $\longrightarrow$ FFFF.......F

By the theorem of compound probability, the probability of getting x successes and $(n - x)$ failures in the above order is

$$\boldsymbol{p.p.p.p\ (x\ times)x\ 9.9\ \ldots\ldots (n - x) times = p^x q^{n-x}}$$

This is one of the orders of getting x successes in n trials. But there can be nC_x different orders of getting x successes out of n trials and for each of these orders the probability will be $\boldsymbol{p^x q^{n-x}}$**.** Hence, by the theorem of addition of probabilities, the probability of getting x successes out of n trials will be,

$$P(x) = nC_x p^x q^{n-x}$$

Above probability function is probability mass function of Binomial Probability distribution. In the above probability mass function $x = 0,1,2, \ldots, n$. Probability of success lies between 0 and 1. (Also note that probability of failure is lies between 0 and 1 & sum of probability of success and failure is 1.)

Binomial distribution can be used for the experiments satisfying the following conditions.

(1) The experiment in which only two results success or failure are obtained.
(2) The experiment can be repeated independently for a fixed number of times n.
(3) The probability of success p remains same in each trial.

9.3.2 Properties of Binomial Probability distribution:

1. This is a distribution of a discrete variable.
2. n and p are the parameters of Binomial distribution.
3. The mean of Binomial distribution is np which shows the average number of successes.
4. The variance of Binomial distribution is npq i.e., its S.D. $= \sqrt{npq}$.
5. When p and q are equal i.e., $p = q = \frac{1}{2}$. Binomial distribution is a symmetrical distribution.
6. When $p < \frac{1}{2}$, its skewness is positive and when $p > \frac{1}{2}$ its skewness is negative.
7. The variance of Binomial distribution is always less than its mean.
8. When number of trials n is very large, and p and q are not very small, Binomial distribution tends to Normal distribution.
9. When number of trials n is very large and p or q is very small, Binomial distribution tends to Poisson distribution.

9.3.3 Uses of Binomial Probability Distribution:

1. For finding out probability of number of successes out of n independent Bernoulli trials.
2. For finding control limits of p and np charts in Statistical Quality Control.
3. In large sample tests for attributes.

Illustration 9.1: An unbiased coin is tossed for 8 times. Find the probabilities of getting (1) 5 heads (2) 8 tails (3) at the most 2 heads.

Answer: Here the probability of getting head in each trial is equal.

Probability of getting head (success) $=\frac{1}{2}$

Therefore, $p=\frac{1}{2}, q=1-p=1-\frac{1}{2}=\frac{1}{2}, n=8$

$$P(x)=nC_x p^x q^{n-x}=8C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{8-x}$$

(1) Probability of getting 5 head out of 8 trial, $(x=5)$

$$P(5)=8C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{8-5}=56\left(\frac{1}{32}\right)\left(\frac{1}{8}\right)=\frac{56}{256}$$

(2) Probability of getting 8 tails that is no head out of 8 trials, $(x=0)$

$$P(0)=8C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{8-0}=1(1)\left(\frac{1}{256}\right)=\frac{1}{256}$$

(3) At the most 2 head, i.e. $(x\leq 2)$

$$P(1)=8C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{8-1}=8\left(\frac{1}{2}\right)\left(\frac{1}{128}\right)=\frac{8}{256}$$

$$P(2)=8C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{8-2}=28\left(\frac{1}{4}\right)\left(\frac{1}{64}\right)=\frac{28}{256}$$

$$P(x\leq 2)=P(0)+P(1)+P(2)$$

$$=\frac{1}{256}+\frac{8}{256}+\frac{28}{256}=\frac{37}{256}$$

Illustration 9.2: The mean of binomial distribution is 20 and variance of binomial distribution is 4. Find the parameter of binomial distribution.

Answer:

Parameter of Binomial Distribution are n and p.

We have given that mean of binomial distribution (np) is 20 and variance of binomial distribution (npq) is 4.

$$Mean = np = 20; \quad Variance = npq = 4$$

To find parameter of Binomial Distribution first of all we divide variance of binomial distribution by mean of binomial distribution,

$$\frac{Variance}{Mean} = \frac{4}{20}$$

$$\frac{npq}{np} = \frac{1}{5}$$

$$q = \frac{1}{5}$$

$$p = 1 - q = 1 - \frac{1}{5}$$

$$p = \frac{4}{5}$$

Here,

$$Mean = 20$$

$$np = 20$$

$$n\left(\frac{1}{5}\right) = 20$$

$$n = 20 \times 5$$

$$n = 100$$

So, parameter of binomial distribution are $n = 100$ and $p = \frac{1}{5}$.

Illustration 9.3: The probability that a person hits a target is $\frac{1}{3}$. Find the probability that he will hit the target in 3 times out of 5 trials.

Answer:

$$\boldsymbol{n = 5, \quad x = 3, \quad p = \frac{1}{3}, \quad q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}}$$

$$P(x) = nC_x p^x q^{n-x} = 5C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{5-x}$$

$$P(3) = 5C_3 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^{5-3} = 10\left(\frac{1}{27}\right)\left(\frac{4}{9}\right)$$

$$P(3) = \frac{40}{243}$$

Illustration 9.4: The probability of a smoker from a group of persons is $\frac{2}{3}$. Five persons are selected at random from the group; find the probability that at least 4 of them are smokers.

Answer:

$$n = 5, \qquad x \geq 4, \qquad p = \frac{2}{3}, \qquad q = \frac{1}{3}$$

$$P(x \geq 4) = P(4) + P(5)$$

$$= 5C_4\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^1 + 5C_5\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^0$$

$$= 5\left(\frac{16}{81}\right)\left(\frac{1}{3}\right) + 1\left(\frac{32}{243}\right)(1)$$

$$= \frac{80}{243} + \frac{32}{243} = \frac{112}{243}$$

The probability that at least 4 of them are smokers $= \frac{112}{243}$.

Illustration 9.5: For a Binomial distribution $n = 6$ and $P(3): P(4) = 8: 3$ find the value of probability of success, mean and Variance of Binomial Distribution.

Answer:

$$P(3): P(4) = 8: 3$$

$$\frac{P(3)}{P(4)} = \frac{8}{3}$$

$$\frac{6C_3\, p^3 q^3}{6C_4\, p^4 q^2} = \frac{8}{3}$$

$$\frac{20q}{15p} = \frac{8}{3}$$

$$60q = 120p$$

$$60\,(1 - p) = 120p$$

$$60 - 60p = 120p$$

$$60 = 60p + 120p$$

$$60 = 180p$$

$$p = \frac{60}{180}$$

$$p = \frac{1}{3}$$

$$Mean = np = 6\left(\frac{1}{3}\right) = 2$$

$$Variance = npq = 6\left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = \frac{12}{9} = \frac{4}{3}$$

Illustration 9.6: The probability that a bomb dropped from a plane will hit a target is $\frac{2}{5}$. Two bombs are enough to destroy a bridge. If 4 bombs are dropped from the plane on a bridge, find the probabilities that (1) The bridge will be saved, (2) the bridge will be

partially destroyed (3) the bridge will be destroyed.

Answer:

Here, $n = 4, \quad p = \frac{2}{5}, \quad q = 1 - p = 1 - \frac{2}{5} = \frac{3}{5}$

(1) The bridge will be saved when there will be no bomb dropped on it i.e. $x = 0$

$$P(0) = 4C_0 \left(\frac{2}{5}\right)^0 \left(\frac{3}{5}\right)^4 = (1)(1)\left(\frac{81}{625}\right) = \frac{81}{625}$$

(2) The bridge will be partially destroyed, when only one bomb dropped on it. (We have said that 2 bombs are enough to destroyed bridge, so in a case when one bomb dropped on a bridge, it will be partially destroyed) $x = 1$

$$P(1) = 4C_1 \left(\frac{2}{5}\right)^1 \left(\frac{3}{5}\right)^3 = 4\left(\frac{2}{5}\right)\left(\frac{27}{125}\right) = \frac{216}{625}$$

(3) The bridge will be destroyed when at least 2 bombs are dropped on it, i.e. $x \geq 2$.

$$P(x \geq 2) = 1 - P(x < 2) = 1 - \{P(0) + P(1)\}$$

$$= 1 - \left\{\frac{81}{625} + \frac{216}{625}\right\} = 1 - \frac{297}{625}$$

$$= \frac{328}{625}$$

9.3.4 Poisson Probability Distribution:

Poisson probability distribution is useful for characterizing events with very low probabilities of occurrence within some definite time or space. Poisson Probability Distribution is a distribution of a discrete random variable, so it is discrete probability distribution. The Poisson distribution was developed by the French mathematician Simeon Denis Poisson in 1837 to describe the number of times a gambler would win a rarely won game of chance in a large number of tries. Letting *p* represent the probability of a win on any given try, the mean (average), number of wins (m) in n tries will be given by m = np. Using the Swiss mathematician Jakob Bernoulli's binomial distribution, Poisson showed that the probability of obtaining k wins is approximately where e is the exponential function, noteworthy is the fact that m equals both the mean and variance (a measure of the dispersal of data away from the mean) for the Poisson distribution.

The Poisson distribution is now recognized as a vitally important distribution in its own right. For example, in 1946 the British statistician R. D. Clarke published "An Application of the Poisson Distribution," in which he disclosed his analysis of the distribution of hits of flying bombs (V-1 and V-2 missiles) in London during World War II. Some areas were hit more often than others. The British military wished to know if the Germans were targeting these districts (the hits indicating great technical precision) or if the distribution was due to chance. If the missiles were in fact only randomly targeted (within a more general area), the British could simply disperse important installations to decrease the likelihood of their being hit.

The Poisson distribution is a discrete distribution that measures the probability of a given number of events happening in a specified time period. In finance, the Poisson distribution could be used to model the arrival of new buy or sell orders entered into the market or the expected arrival of orders at specified trading venues or dark pools. In these cases, the Poisson distribution is used to provide expectations surrounding confidence bounds around the expected order arrival rates. Poisson distributions are very useful for smart order routers and algorithmic trading.

The Poisson random variable satisfies the following conditions:

1. The number of successes in two disjoint time intervals is independent.
2. The probability of a success during a small time interval is proportional to the entire length of the time interval.

Apart from disjoint time intervals, the Poisson random variable also applies to disjoint regions of space.

The **probability distribution of a Poisson random variable** X representing the number of successes occurring in a given time interval or a specified region of space is given by the formula:

$$P(x) = \frac{e^{-m}m^x}{x!}$$

Where,

$x = 0, 1, 2, 3, \dots (number\ of\ success)$

$e = 2.71828\ (constant)$

$m =$ mean number of successes in the given time interval or region of space

9.3.4 Properties of Poisson Probability distribution:

(1) This is a distribution of a discrete variable.
(2) It is a distribution of rare occurrence.
(3) m is parameter of the distribution.
(4) The mean of this distribution is m.
(5) The variance of this distribution is also m. Hence, its S.D. = $\sqrt{m}$. i.e., in Poisson distribution mean = variance.
(6) Sum of two independent Poisson variates is also a Poisson variate.
(7) This is a distribution with positive skewness.

9.3.5 Application of Poisson Probability Distribution:

The following are some of the instances where Poisson distribution can be applied:

(1) Number of accidents on a road.
(2) Number of misprints per page of a book.
(3) Number of defects in a radio set.
(4) Number of air bubbles in a glass bottle.
(5) Number of suicides committed per day.
(6) Number of goals scored in a football match.
(7) Number of telephone calls received during a given interval of time.
(8) The number of deaths by horse kicking in the Prussian army (first application)

(9) Birth defects and genetic mutations

(10) Rare diseases (like Leukemia, but not AIDS because it is infectious andso not independent) - especially in legal cases

Thus, Poisson distribution is a distribution of rare events.

Illustration 9.7: A person has some cars and average demand of cars per day is 3, find the probability that on any day (1) only one car is in use (2) not more than 2 cars are in use. ($e^{-3} = 0.0498$)

Answer: Here we have given that average demand of cars per day is 3 i.e. m = 3.

$$P(x) = \frac{e^{-m}m^x}{x!} = \frac{e^{-3}3^x}{x!} = \frac{(0.0498)(3^x)}{x!}$$

(1) Only one car is in use, therefore putting $x = 1$

$$(1) = \frac{0.0498\ (3^1)}{1!} = \frac{0.0498(3)}{1} = 0.1494$$

(2) Not more than two cars are in use, i.e. $x \leq 2$

$$(x \leq 2) = (0) + (1) + (2) = \frac{0.0498\ (3^0)}{0!} + \frac{0.0498\ (3^1)}{1!} + \frac{0.0498\ (3^2)}{2!}$$

$$= \frac{0.0498(1)}{1} + \frac{0.0498(3)}{1} + \frac{0.0498(9)}{2} = 0.0498 + 0.1494 + 0.2241$$

$$(x \leq 2) = 0.4233$$

Illustration 9.8: in the production of electric fuses 2% are defective. Find the probability of getting (i) all non-defective fuses (ii) at the most 2 defective fuses (iii) 3 defective fuses in a box containing 200 fuses. ($e^{-4} = 0.0183$)

Answer: Here, sample size n is large and probability of success is too small so we use Poisson Probability Distribution to calculate desire probabilities.

Average defective fuse in sample of 200 fuses = 200 (2%) = 200 (0.02) = 4 i.e. $m = 4$

$$P(x) = \frac{e^{-m}m^x}{x!} = \frac{e^{-4}4^x}{x!} = \frac{(0.0183)(4^x)}{x!}$$

(i) All non-defective fuses i.e. $x = 0$

$$(0) = \frac{(0.0183)(4^x)}{x!} = \frac{(0.0183)(4^0)}{0!} = \frac{(0.0183)(1)}{(1)} = 0.0183$$

(ii) At the most 2 defective fuses i.e. $x \leq 2$

$$(x \leq 2) = (0) + (1) + (2) = \frac{0.0183\ (4^0)}{0!} + \frac{0.0183\ (4^1)}{1!} + \frac{0.0183\ (4^2)}{2!}$$

$$= \frac{0.0183(1)}{1} + \frac{0.0183(4)}{1} + \frac{0.0183(16)}{2} = 0.0183 + 0.0732 + 0.1464$$

$$(x \leq 2) = 0.2379$$

(iii) 3 defective fuses i.e. $x = 3$

$$(3) = \frac{0.0183\ (4^3)}{3!} = \frac{0.0183(64)}{6} = 0.1952$$

Illustration 9.9: In the manufacturing of cotter pins, it is known that 5% of the pins are defective. The pins are sold in boxes of 100 and it is guaranteed that not more than 4 pins will be defective in a box. What is the probability that a box will not meet this guarantee? ($e^{-5} = 0.0067$)

Answer: As we discuss earlier when n is large and probability of success is small then we use Poisson probability distribution. To use Poisson probability distribution we need to find the parameter of Poisson Probability Distribution.

$$m = np = 100(5\%) = 100(0.05) = 5$$

$$P(x) = \frac{e^{-m}m^x}{x!} = \frac{e^{-5}5^x}{x!} = \frac{0.0067(5^x)}{x!}$$

To find the probability of box will not meet given guarantee we have to calculate that more than 4 pins are defective. i.e. $x > 4$

$$(x > 4) = (5) + (6) + \cdots + (100)$$

$$= 1 - \{(0) + (1) + (2) + (3) + (4)\}$$

$$= 1 - \{\frac{0.0067(5^0)}{0!} + \frac{0.0067(5^1)}{1!} + \frac{0.0067(5^2)}{2!} + \frac{0.0067(5^3)}{3!} + \frac{0.0067(5^4)}{4!}\}$$

$$= 1 - \{0.0067 + 0.0335 + 0.0838 + 0.1396 + 0.1745\}$$

$$= 1 - 0.4381 = 0.5619$$

So, probability that a box will not meet this guarantee is **0.5619**.

Illustration 9.10: For a Poisson Variate x, $(x = 1) = (x = 2)$. Find mean, variance, standard deviation and (4).

Answer: We have given that,

$$P(x = 1) = P(x = 2)$$

$$\frac{e^{-m}m^1}{1!} = \frac{e^{-m}m^2}{2!}$$

$$\frac{2!}{1!} = \frac{e^{-m}m^2}{e^{-m}m^1}$$

$$m = 2$$

Mean of Poisson Distribution $(m) = 2$

Variance of Poisson Distribution $(m) = 2$

Standard Deviation of Poisson Distribution $(\sqrt{m}) = \sqrt{2} = 1.4142$

$$\boldsymbol{P(4)} = \frac{e^{-2}2^4}{4!} = \frac{0.1353(16)}{24} = 0.0902$$

Illustration 9.11: The mean of a Poisson Variate is 0.81, find its S.D. and the probabilities for $x = 0$ and $x = 2$.

Answer: Mean of a Poisson Variate $m = 0.81$

Standard Deviation of Poisson Distribution $(\sqrt{m}) = \sqrt{0.81} = 0.9$

$$(x = 0) = \frac{e^{-0.81}(0.81^0)}{0!} = \frac{0.4449(1)}{1} = 0.4449$$

$$(x = 2) = \frac{e^{-0.81}(0.81^2)}{2!} = \frac{0.4449(0.6561)}{2} = 0.1459$$

9.3.6 Hypergeometric Probability Distribution:

The hypergeometric distribution occupies a place of great significance in statistical theory. It applies to sampling without replacement from a finite population whose elements can be classified into two categories-one which possesses a certain characteristic and another which does not possess that characteristic. The categories could be-male, female, employed, unemployed, etc. When random selections are made without replacement from the population, each subsequent draw is dependent and the probability of success changes in each draw. The following conditions characterize the hypergeometric distribution:

1. The result of each draw can be classified into one of two categories;
2. The probability of a success changes on each draw;
3. Successive draws are dependent; and
4. The drawing is repeated a fixed number of times

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 hypergeometric probability distribution which gives the probability of r successes is given by

$$P(r) = \frac{({}^{k}C_{r})({}^{N-k}C_{n-r})}{{}^{N}C_{n}}$$

Here, r is a discrete random variable which can take values 0, 1, 2,…, n. Also $n \leq k$

Both the binomial distribution and the hypergeometric distribution are concerned with the same thing, viz., number of successes in a sample containing n observations. What differentiates these two discrete probability distributions is the manner in which data are obtained. For the binomial model, the sample data are drawn with replacement from a finite population or without replacement from an infinite population. On the other hand, for the hypergeometric model, the sample data are drawn without replacement from a finite population. Hence, while the probability of success p is constant over all observations of a binomial experiment and the outcome of any particular observation is independent of any other, the same cannot be said for the hypergeometric experiment; here the outcome of one observation is affected by the outcomes of the previous observations.

The hypergeometric distribution bears a very interesting relationship to the binomial distribution. When N increases without limit, the hypergeometric distribution approaches the binomial distribution. Hence, the binomial probabilities may be used as approximation to hypergeometric probabilities where σ/N is small. A frequently used rule of thumb is that the population size should be at least ten times the sample size (N> 10n) for the approximation to be used.

9.3.7 Properties of Hypergeometric Probability Distribution

A hypergeometric experiment is a statistical experiment with the following properties:

(1) You take samples from two groups.

(2) You are concerned with a group of interest, called the first group.

(3) You sample without replacement from the combined groups.

(4) Each pick is not independent, since sampling is without replacement.

(5) Mean of the hyper geometric distribution is $\frac{nk}{N}$.

(6) Variance of the hyper geometric distribution is $\frac{(N-n)(N-k)}{N^2(N-1)}$.

Illustration 9.12: A bag contains 20 balls of which 15 are of red colour and 5 are of black colour. A random sample (without replacement) of 5 balls is taken. Find the probability that the sample contains 3 black balls.

Answer:

The problem can be solved with the help of hypergeometric probability model.

Here, N = 20, k = 15, n = 5, r = 3

$$(2) = \frac{({}^{5}C_3)({}^{20-5}C_{5-2})}{{}^{20}C_5} = \frac{(10)(455)}{15504} = 0.29$$

Illustration 9.13: A retailer has 10 identical television sets of a company out which 4 are defective. If 3 televisions are selected at random, find the probability that selected (i) at least 1 defective television is defective (ii) only 2 televisions are defective (iii) no television are defective.

Answer:

Here, N = 10, k = 4, n = 3.

$$P(r) = \frac{({}^{k}C_r)({}^{N-k}C_{n-r})}{{}^{N}C_n}$$

(i) At least 1 defective television is selected i.e. $x \geq 1$

$$(x \geq 1) = 1 - (x < 1) = 1 - (0)$$

$$= 1 - \frac{({}^{4}C_0)({}^{10-4}C_{3-0})}{{}^{10}C_3} = 1 - \frac{(1)(20)}{120} = 1 - \frac{1}{6} = \frac{5}{6}$$

(ii) Only 2 televisions are defective, i.e. $x = 2$

$$(2) = \frac{({}^{4}C_2)({}^{10-4}C_{3-2})}{{}^{10}C_3} = \frac{(6)(6)}{120} = \frac{6}{20} = \frac{3}{10} = 0.3$$

(iii) No television are defective, i.e. $x = 0$

$$(0) = \frac{({}^{4}C_0)({}^{10-4}C_{3-0})}{{}^{10}C_3} = \frac{(1)(20)}{120} = \frac{1}{6}$$

9.4 CONTINUOUS PROBABILITY DISTRIBUTION

We discussed the probability distribution of the discrete random variable. Further we wish to study probability distribution of a continuous random variable. A continuous distribution has a range of values that are infinite, and therefore uncountable. For example, time is infinite: you could count from 0 seconds to a billion seconds…a trillion seconds…and so on, forever. A discrete distribution has a range of values that are countable. For example, the numbers on birthday cards have a possible range from 0 to 122 (122 is the age of Jeanne Calment the oldest person who ever lived). Such kind of variable has probability distribution which is known as probability density function or continuous probability distribution. Some important continuous probability distributions are listed below. We learn about it.

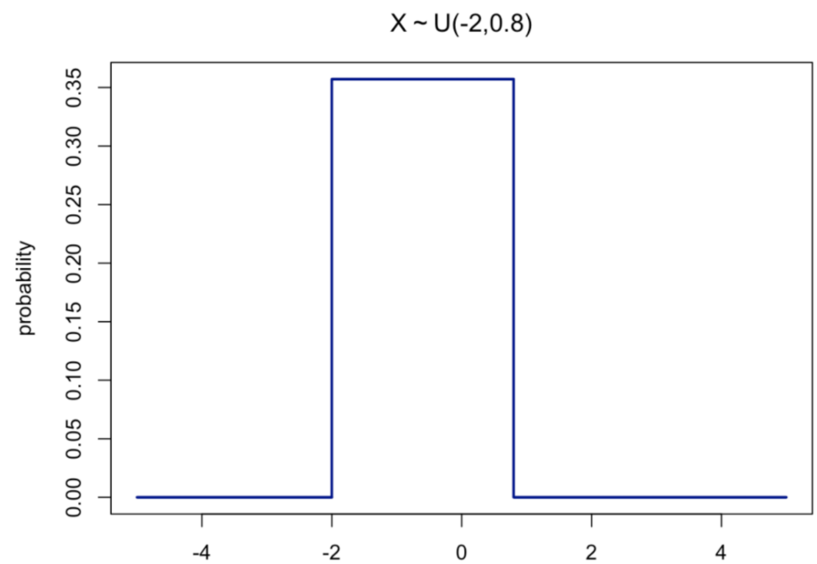
The probability distribution that deals with continuous random variable is called the probability density function (pdf). There are various types of continuous probability distribution. The normal distribution is the "go to" distribution for many reasons, including that it can be used the approximate the binomial distribution, as well as the hypergeometric distribution and Poisson distribution. Other continuous distributions that are common in statistics include: Uniform Distribution, Beta distribution, Cauchy distribution, Exponential distribution, Gamma distribution, Logistic distribution, Weibull distribution.

9.4.1 Uniform Probability Distribution:

The Uniform distribution is the simplest probability distribution, but it plays an important role in statistics since it is very useful in modeling random variables. The uniform distribution is a continuous probability distribution and is concerned with events that are equally likely to occur. The continuous random variable X is said to be uniformly distributed, or having rectangular distribution on the interval [a, b]. We write X~U(a, b), if its probability density function equals

$$f(x) = \begin{cases} \frac{1}{b-a}; & when\ a \le x \le b \\ 0; & when\ x < a\ or\ x > b \end{cases}$$

The figure below shows a continuous uniform distribution $X \sim (-2, 0.8)$, thus a distribution where all values of x within the interval [-2, 0.8] are $\frac{1}{b-a}$ $(= \frac{1}{0.8-(-2)} = 0.36)$, whereas all other values of x are 0.



9.4.2 Properties of Uniform Probability Distribution:

The following are the key properties of the Uniform Distribution:

- The density function integrates to unity
- Each of the points that go into form the function have equal weighting
- Mean of the Uniform Distribution is $\frac{(a+b)}{2}$
- The variance of the Uniform Distribution is $\frac{(b-a)^2}{12}$

9.4.3 Application of Uniform Probability Distribution:

The probabilities for uniform distribution function are simple to calculate due to the simplicity of the function form. Therefore, there are various applications that this distribution can be used for as shown below: hypothesis testing situations, random sampling cases, finance, etc. Furthermore, generally, experiments of physical origin follow a uniform distribution (e.g. emission of radioactive particles). However, it is important to note that in any application, there is the unchanging assumption that the probability of falling in an interval of fixed length is constant.

- One of the most important applications of the uniform distribution is in the generation of random numbers. That is, almost all random number generators generate random numbers on the (0, 1) interval.
- In the field of economics, usually demand and replenishment may not follow the expected normal distribution. As a result, other distribution models are used to better predict probabilities and trends such as Bernoulli process. But according to Wanke (2008), in the particular case of investigating lead-time for inventory management at the beginning of the life cycle when a completely new product is being analyzed, the uniform distribution proves to be more useful. In this situation, other distribution may not be viable since there is no existing data on the new product or that the demand history is unavailable so there isn't really an appropriate or known distribution. The uniform distribution would be ideal in this situation since the random variable of lead-time (related to demand) is unknown for the new product but the results are likely to range between a plausible range of two values. The lead-time would thus represent the random variable. From the uniform distribution model, other factors related to lead-time were able to be calculated such as cycle service level and shortage per cycle. It was also noted that the uniform distribution was also used due to the simplicity of the calculations.
- Guessing a Birthday, if you randomly approach a person and try to guess his/her birthday, the probability of his/her birthday falling exactly on the date you have guessed follows a uniform distribution. This is because every day of the year has equal chances of being his/her birthday or every day of the year is equally likely to be his/her birthday. For instance, the probability that the 1st of January is supposed to be his/her birthday is equal to 1/365, which is the same as the probability that the 2nd of January is his/her birthday, which is the same as the probability of each and every day of the year to be his/her birthday.
- Rolling a Dice, when a fair die is rolled, the probability that the number appearing on the top of the die lies in between one to six follows a uniform distribution. The probability that number ‘one’ will appear on the top of the die is equal to 1/6, which is the same as the probability that number ‘two’ will appear on the top of

the dice, and so on. Each number has equal chances of occurring on the top, hence the distribution is uniform in nature.

- Tossing a Coin, when you flip a coin, the probability of the coin landing with a head faced up is equal to the probability that it lands with a tail faced up. Since the experiment of tossing the coin has two outcomes each of which is equally likely to occur, it is said to be following a uniform distribution.
- Deck of Cards, the total number of cards present in the deck of playing cards is equal to 52. The deck is further divided into four sets of 13 cards each of which is marked with certain shapes including diamond, spade, heart, and club. If you select a card randomly from a fair deck of playing cards, the probability that the drawn card would be either a diamond, spade, heart, or club follows a uniform distribution because the probability of choosing a spade is equal to 0.25, which is same as the probability of choosing a diamond, heart, or a club card.
- Spinning a Spinner, suppose a spinner is rotated over a tray that consists of four individual compartments. Each compartment is coloured different from others. The probability that the spinner after spinning would point towards any of the four coloured compartments is equal to 0.25. Hence, it forms a prominent example of uniform distribution in real life because each colour compartment has equal chances to be pointed by the spinner.
- Raffle Tickets, a raffle is a prominent example of a uniform probability distribution. The event's organizing committee tends to select a particular seat out of thousands of seats and reward the person sitting on it with a prize. The people participating in the event buy numbered raffle tickets and each of them possesses an equal chance of winning. This is because the probability of a seat being chosen by the organizers as the winner is equal to the remaining seats.
- Lucky Draw Contest, the probability of a person winning a lucky draw contest is equal for every other person participating in the contest. Hence, such a distribution is known as the uniform probability distribution because the winning chances of every person are equal.

- Throwing a Dart, when you throw a dart at the dartboard, each and every point of the dartboard has an equal probability of getting hit by it. Hence, it is a prime example of uniform distribution in real life.

Illustration 9.14: The amount of gasoline sold every day at a service station is uniformly distributed. The minimum sold is 1000 gallons and maximum sold is 3000 gallons. Find the probability that the service station will sell at least 2,000 gallons.

Answer:

Here, we wish to calculate the probability that at least 2000 gallons will be sell on a particular day that means more or equal to 2000 gallons will be sell (maximum 3000 gallons will be on a particular day that is given). In this example maximum value (b) is 3000 and minimum value is (a) is 1000.

$$(x \geq 200) = \int_{2000}^{3000} f(x)dx = \int_{2000}^{3000} \frac{1}{b-a} dx = [\frac{x}{b-a}]_{2000}^{3000}$$

$$= [\frac{3000-2000}{3000-1000}] = \frac{1000}{2000} = 0.5$$

Illustration 9.15: You arrive into a building and are about to take an elevator to the floor. Once you call the elevator, it will take between 0 and 40 seconds to arrive to you. We will assume that the elevator arrives uniformly between 0 and 40 seconds after you press the button. Find the probability that (i) you have to wait 15 seconds for the elevator (ii) you have to wait 20 seconds for the elevator.

Answer: In this example a = 0 and b = 40.

$$\boldsymbol{f(x) = \frac{1}{b-a} = \frac{1}{40-0}}$$

(i) You have to wait 15 seconds for the elevator.

$$\boldsymbol{P(0 \leq x \leq 15)} = \int_{0}^{15} f(x)dx = \int_{0}^{15} \frac{1}{40-0} dx = \frac{15-0}{40-0} = \frac{15}{40} = 0.375$$

(ii) You have to wait 20 seconds for the elevator.

$$\boldsymbol{P(0 \leq x \leq 20)} = \int_{0}^{20} f(x)dx = \int_{0}^{20} \frac{1}{40-0}dx = \frac{20-0}{40-0} = 0.5$$

9.4.4 Normal Probability Distribution

There are many distribution of continuous variable, but the most important of them is **Normal Distribution**. Normal distribution has a wide application in statistics. This distribution was first given by De Movire in 1773. He developed this distribution as a limiting case of Binomial distribution. In Binomial distribution when n is very large and p and q are not very small, it tends to Normal distribution. **Gauss** and **Laplace** also derived the mathematical form of this distribution from the study of the distribution of errors. The graph of normal distribution called normal curve is bell-shaped and it extends indefinitely on both the sides of the x-axis but does not meet it i.e. the behavior of the tails of the curve is asymptotic to the x-axis.

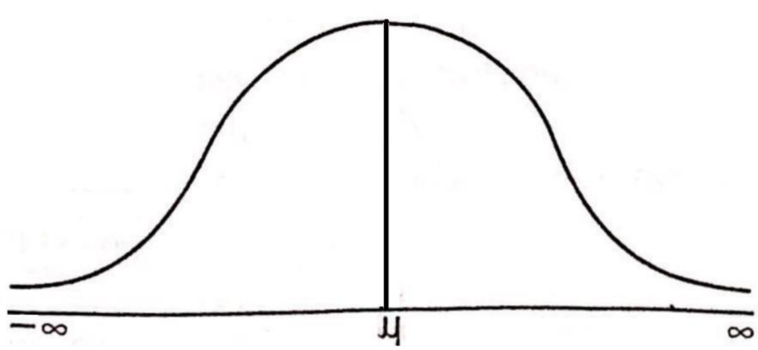
A continuous random variable X is normally distributed or follows a normal probability distribution if its probability distribution is given by the following function:

$$\boldsymbol{f(x)} = \frac{\mathbf{1}}{\boldsymbol{\sigma\sqrt{2\pi}}}\boldsymbol{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2};$$

$$-\infty < x < \infty, \qquad -\infty < \mu < \infty, \qquad 0 < \sigma^2 < \infty$$

Here μ and σ are mean and standard deviation of the normal distribution and they are called the parameters of the normal distribution whereas $e = 2.7183$ and $\pi = \frac{22}{7} = 3.14$ are numerical constants.

We saw the probability density function of the normal distribution for random variable x with parameters μ and σ. If parameters are known we can draw a probability curve of the distribution. This curve is known as a normal cure.



A normal curve is symmetrical on both the sides of the central line. The area between the normal curve and x axis is taken as 1, and also the total probability of a variable x taking values between -∞ and +∞ is 1. So the probability that the values of x will lie between -∞ and ∞ = The total area under the curve = 1.

Moreover, the probability that a random variable x will assume value between a and b can be given by the area under the curve between the ordinates at $x = a$ and $x = b$ i.e. $Pr \quad b. (a \leq x \leq b)$ = Area under the curve between the ordinates at a and b. Thus, in order to find the probability that the value of x lies between a and b, we should find the area under the curve between a and b.

We have discussed that normal distribution can be completely specified by its mean μ and standard deviation σ. For different values of mean μ and standard deviation σ, we can draw different normal curves and from these curves we can find the probability of a value of x to lie within a certain interval. These probabilities can be represented in tables. For different pairs of values of μ and σ, we may be required to prepare many such tables but the difficulty is solved by converting a variable x into a variable Z with mean 0 and S.D. 1. The variable Z is known as a standard normal variate. For different values of Z the probability tables are prepared. The probability function of standard normal variate Z can be given as follows:

$$\boldsymbol{f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}};$$

$$-\infty < \mathrm{Z} < \infty, \qquad \mathrm{Z} = \frac{x -}{\sigma}, \qquad \mu = 0, \qquad \sigma^2 = 1$$

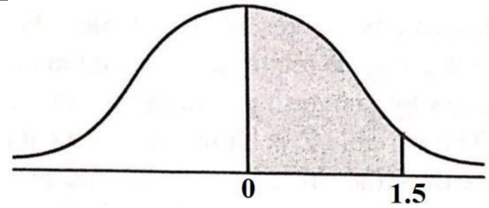
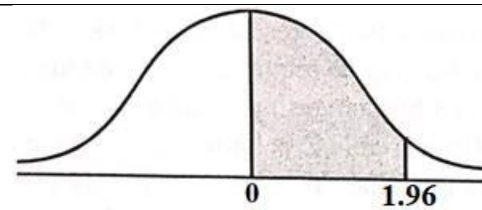
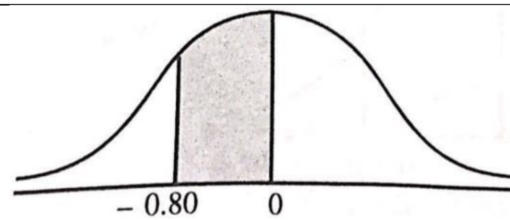
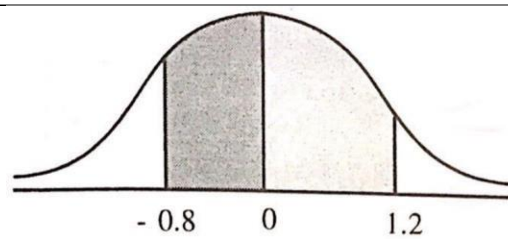
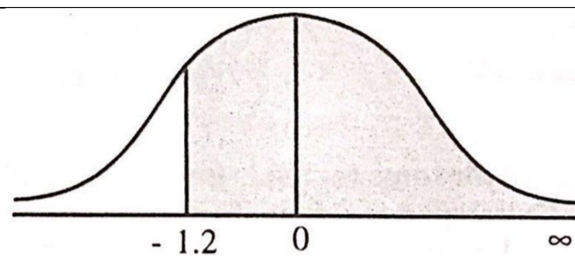
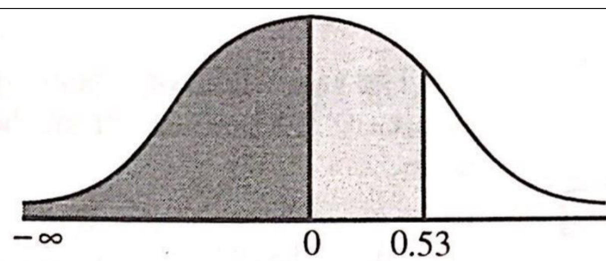
The probability curve of a standard normal variate is known as a standard normal curve. The total area under the normal curve is 1 and it is a symmetrical curve.

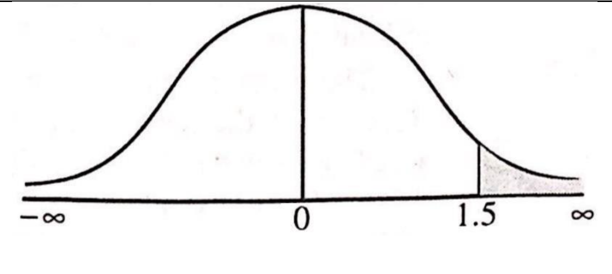
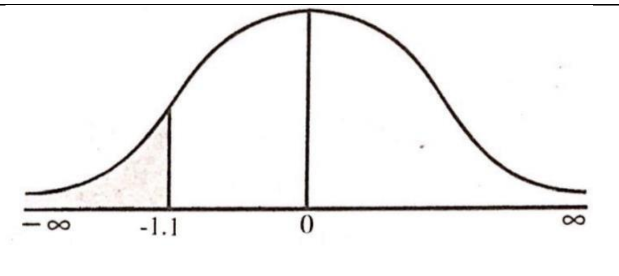
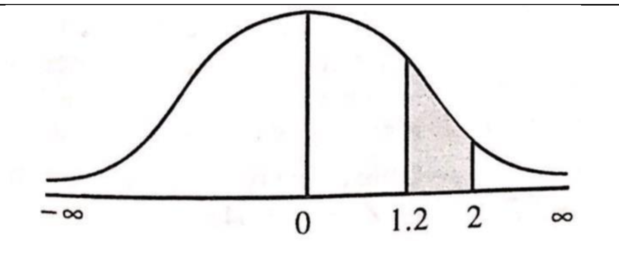
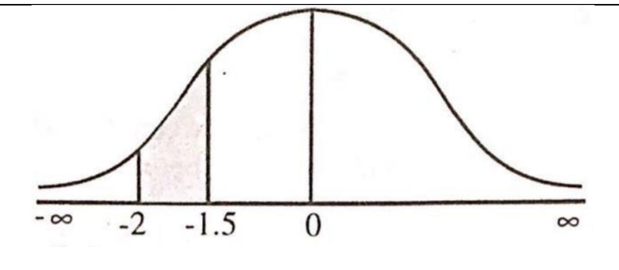
The area of the curve for the values of Z between -∞ and 0 = Area of the curve for the values of Z between 0 and ∞ = 0.5.

Tables are prepared giving areas of the normal curve between Z=0 and different positive values of Z. From these tables, the area between Z = 0 and Z = some value Z_1 can be found out. This area gives the probability of the value of Z to lie between Z = 0 and Z = Z_1.

Thus, from the tables we can find out the area between any two values of standard normal variate. This area is the probability that Z will lie between these two values. For example the area between Z = 0 and Z = 1 is 0.3413 from the table. If we want area between

Z= -1 and Z = 0, we should refer the table for finding the area between Z = 0 and Z = + 1. The curve, being symmetrical the area between -1 and 0 = the area between 0 and + 1 = 0.3413. Now we shall find area between any two values of Z from the table.

(1) Probability that the value of Z lies between 0 and 1.5 $P\{0 \leq Z \leq 1.5\} = 0.4332$	
(2) Probability that the value of Z lies between 0 and 1.96 $P\{0 \leq Z \leq 1.96\} = 0.4750$	
(3) Probability that the value of Z lies between -0.80 and 0 $P\{-0.80 \leq Z \leq 0\} = 0.2881$	
(4) Probability that the value of Z lies between -0.80 and 1.2 $P\{-0.80 \leq Z \leq 1.2\} =$ $P\{-0.80 \leq Z \leq 0\} + P\{0 \leq Z \leq 1.2\}$ $= 0.2881 + 0.3849 = 0.6730$	
(5) Probability that the value of Z is more than -1.2 $P\{Z \geq -1.2\} = P\{-1.2 \leq Z \leq \infty\}$ $= P\{-1.2 \leq Z \leq 0\} + P\{0 \leq Z \leq \infty\}$ $= 0.3849 + 0.5 = 0.8849$	
(6) Probability that the value of Z is less than 0.53 $P\{Z \leq 0.53\} = P\{-\infty \leq Z \leq 0.53\}$ $= P\{-\infty \leq Z \leq 0\} + P\{0 \leq Z \leq 0.53\}$ $= 0.5 + 0.2019 = 0.7019$	

(7) Probability that the value of Z is more than 1.5 $P\{Z \geq 1.5\} = P\{1.5 \leq Z \leq \infty\}$ $= P\{0 \leq Z \leq \infty\} - P\{0 \leq Z \leq 1.5\}$ $= 0.5 - 0.4332 = 0.0668$	
(8) Probability that the value of Z is less than -1.1 $P\{Z \leq -1.1\} = P\{-\infty \leq Z \leq -1.1\}$ $= P\{-\infty \leq Z \leq 0\} - P\{0 \leq Z \leq -1.1\}$ $= 0.5 - 0.3643 = 0.1357$	
(9) Probability that the value of Z lies between 1.2 and 2 $P\{1.2 \leq Z \leq 2\} =$ $P\{0 \leq Z \leq 2\} - P\{0 \leq Z \leq 1.2\}$ $= 0.4772 - 0.3849 = 0.0923$	
(10) Probability that the value of Z lies between -2 and -1.5 $P\{-2 \leq Z \leq -1.5\} =$ $P\{-2 \leq Z \leq 0\} - P\{-1.5 \leq Z \leq 0\}$ $= 0.4772 - 0.4332 = 0.0440$	

9.4.5 Properties of Normal Probability Distribution

(1) This is a distribution of a continuous variable.

(2) μ and σ are the parameters of this distribution.

(3) The curve of the normal distribution is symmetrical about the mean and it is bell-shaped.

(4) Mean, median and mode are equal in this distribution.

(5) Quartiles are equidistant from median.

(6) Its skewness is zero.

(7) The total area under the normal curve is 1.

The following are some of the important areasof the normal curve:

(a) Area between $\mu \pm \sigma = 0.6826$

(b) Area between $\mu \pm 2\sigma = 0.9545$

(c) Area between $\mu \pm 3\sigma = 0.9973$

(d) Area between $\mu \pm 1.96\sigma = 0.95$

(e) Area between $\mu \pm 2.58\sigma = 0.99$

(8) The tails of the normal curve do not meet x axis. i.e., the curve is asymptotic to the x-axis.

(9) Mean deviation about mean= $\frac{4}{5}\sigma$ (approximately)

(10) Quartile deviation = $\frac{2}{3}\sigma$ (approximately)

(11) The sum of two independent normal variates is also a normal variate.

(12) When n is very large, and p and q are not very small, Binomial distribution tends to Normal distribution.

(13) In fact, most of the distributions in statistics follow normal distribution when n is large.

9.4.6 Importance of Normal Distribution

The most important continuous probability distribution widely used in statistics is normal distribution. Its importance can be given as follows:

(1) Most of the distribution like Binomial, Poisson tend to Normal distribution when n is very large.

(2) In Practice most of the events follow normal distribution. Hence, it is used in the fields of Sociology, Psychology, Economics, etc.

(3) In statistical quality control, the properties of normal distribution are used.

(4) In significant tests for large samples, the distribution is widely used.

(5) It is used in small sample tests in which it is assumed that the population from which the sample is drawn is normal.

Illustration 9.16: In an intelligence test administered of children the average score is 42 and its s.d. is 24. Assuming that the scores are normally distributed find the probability that (i) children exceeding score 50. (ii) Children getting score between 30 and 54.

Answer: Here, in this illustration $\mu = 42$ and $\sigma = 24$

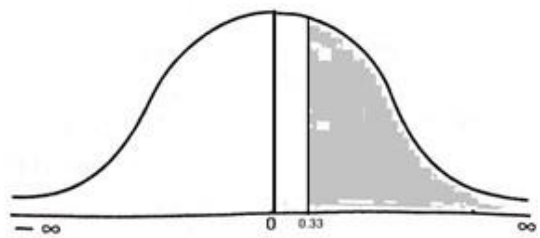
(i) Children exceeding score 50 i.e. $x \geq 50$

$$P(x \geq 50) = P\left(\frac{x-\mu}{\sigma} \geq \frac{50-\mu}{\sigma}\right)$$

$$= P\left(Z \geq \frac{50-42}{24}\right)$$

$$= P\left(Z \geq \frac{8}{24}\right) = P(Z \geq 0.33)$$

$$= P(0 < Z < \infty) - P(0 < Z \leq 0.33)$$



$$= 0.5 - 0.1293 = 0.2707$$

(ii) Children getting score between 30 and 54

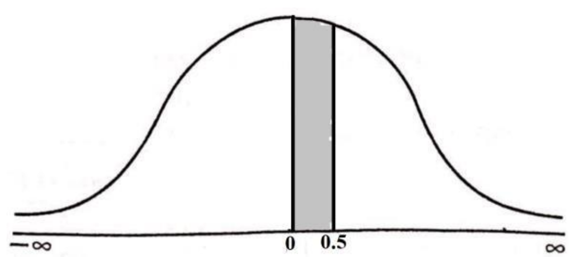
$$P(30 \leq x \leq 54)$$

$$= P\left(\frac{30-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{54-\mu}{\sigma}\right)$$

$$= P\left(\frac{30-42}{24} \leq Z \leq \frac{54-42}{24}\right)$$

$$= P(-0.5 \leq Z \leq 0.5)$$

$$= P(-0.5 \leq Z \leq 0) + P(0 \leq Z \leq 0.5)$$



$$= 2 \times P(0 \leq Z \leq 0.5) = 2(0.1915)$$

$$= 0.3830$$

Illustration 9.17: Average salary of employee is 750 and s.d. of salary is 50. Find the probability that employee has salary (i) less than 650 (ii) more than 800 and (iii) between 700 to 850.

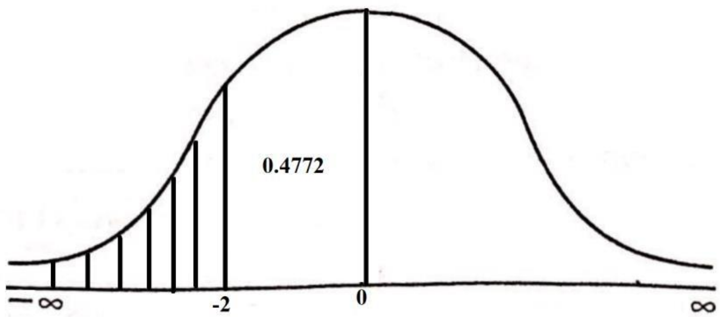
Answer:

$$\boldsymbol{\mu = 750; \quad \sigma = 50}$$

(i) salary less than 650 i.e. $x < 650$

$$P(x < 650) = P\left(\frac{x-\mu}{\sigma} < \frac{650-\mu}{\sigma}\right) P(Z < -2) = 0.5 - P(-2 < Z < 0)$$

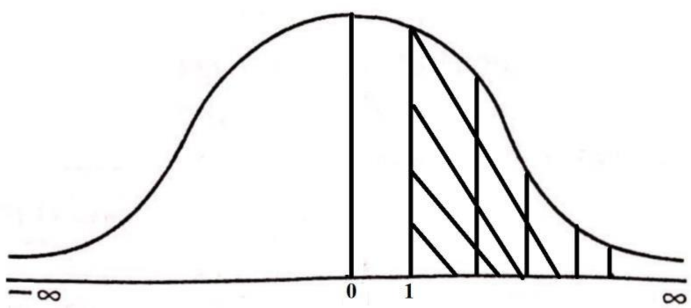
$$= 0.5 - 0.4772 = 0.0228$$



(ii) Salary less than 800 i.e. $x < 800$

$$P(x > 8\quad 0) = P\left(\frac{x-\mu}{\sigma} > \frac{800-\mu}{\sigma}\right) = P(Z > 1) = 0.5 - P(0 < Z < 1)$$

$$= 0.5 - 0.3413 = 0.1587$$



(iii) Salary between 700 to 850 i.e. $700 < x < 850$

$$P(700 < x < 850) = P\left(\frac{700-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{850-\mu}{\sigma}\right) = P(-1 < Z < 2)$$

$$= 0.3413 + 0.4772 = 0.8185$$

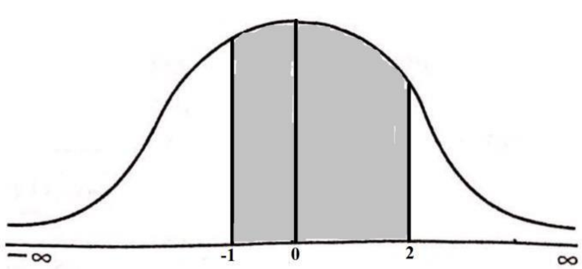


Illustration 9.18: For a normal probability distribution of 100 items Q_1 = 73 and σ = 15, find mean, median, mode and limits for central 50% of the items.

Answer:

For a quartile Deviation of Normal Distribution $= \frac{2}{3}\sigma = \frac{2}{3}(15) = 10$

Quartile Deviation = 10

$$\frac{Q_3 - Q_1}{2} = 10$$

$$\frac{73 - Q_1}{2} = 10$$

$$73 - Q_1 = 20$$

$$Q_1 = 73 + 20 = 93$$

For a normal distribution median

$$= \frac{Q_3 + Q_1}{2} = \frac{73 + 93}{2} = 83$$

For a normal distribution Mean = Median = Mode = 83

The limits of central 50% of the items is Q_1 and Q_3 i.e. 73 to 93.

❖ **CHECK YOUR PROGRESS**

- **Answer the following Multiple Choice Questions**

Que. 1 Which of the following mentioned standard Probability density functions is applicable to discrete Random Variables?
(a) Gaussian Distribution **(b)** Poisson Distribution **(c)** Rayleigh Distribution **(d)** Exponential Distribution

Que. 2 A variable that can assume any value between two given points is called ___________
(a) Continuous random variable **(b)** Discrete random variable **(c)** Irregular random variable **(d)** Uncertain random variable

Que. 3 If f(x) is a probability density function of a continuous random variable, then $\int_{-\infty}^{\infty} (x)$=?
(a) 0 **(b)** 1 **(c)** undefined **(d)** Insufficient data

Que. 4 A random variable that assumes a finite or a countably infinite number of values is called ___________
a) Continuous random variable
b) Discrete random variable
c) Irregular random variable
d) Uncertain random variable

Que. 5 In a discrete probability distribution, the sum of all probabilities is always?
(a) 0 (b) Infinite (c) 1 (d) Undefined

Que. 6 What is the area under a conditional Cumulative density function?
(a) 0 **(b)** Infinity **(c)** 1 **(d)** Changes with CDF

Que. 7 In a Binomial Distribution, if 'n' is the number of trials and 'p' is the probability of success, then the mean value is given by ___________
a) np
b) n
c) p
d) np(1-p)

Que. 8 In a Binomial Distribution, if p, q and n are probability of success, failure and number of trials respectively then variance is given by ___________
a) np
b) npq
c) np^2q
d) npq^2

Que. 9 If 'X' is a random variable, taking values 'x', probability of success and failure being 'p' and 'q' respectively and 'n' trials being conducted, then what is the probability that 'X' takes values 'x'? Use Binomial Distribution
a) $P(X = x) = {}^{n}C_{x}\, p^{x}\, q^{x}$
b) $P(X = x) = {}^{n}C_{x}\, p^{x}\, q^{(n-x)}$
c) $P(X = x) = {}^{x}C_{n}\, q^{x}\, p^{(n-x)}$
d) $P(x = x) = {}^{x}C_{n}\, p^{n}\, q^{x}$

Que. 10 If ‘p’, ‘q’ and ‘n’ are probability pf success, failure and number of trials respectively in a Binomial Distribution, what is its Standard Deviation?

a) $\sqrt{np}$
b) $\sqrt{pq}$
c) $(np)^2$
d) $\sqrt{npq}$

Que. 11 In a Binomial Distribution, the mean and variance are equal.

a) True
b) False

Que. 12 It is suitable to use Binomial Distribution only for ____________

a) Large values of ‘n’
b) Fractional values of ‘n’
c) Small values of ‘n’
d) Any value of ‘n’

Que. 13 For larger values of ‘n’, Binomial Distribution ____________

a) loses its discreteness
b) tends to Poisson Distribution
c) stays as it is
d) gives oscillatory values

Que. 14 In a Binomial Distribution, if p = q, then P(X = x) is given by?

a) ${}^nC_x\ (0.5)^n$
b) ${}^nC_n\ (0.5)^n$
c) ${}^nC_x\ p^{(n-x)}$
d) ${}^nC_n\ p^{(n-x)}$

Que. 15 Binomial Distribution is a ____________

a) Continuous distribution
b) Discrete distribution
c) Irregular distribution
d) Not a Probability distribution

Que. 16 In a Poisson Distribution, if ‘n’ is the number of trials and ‘p’ is the probability of success, then the mean value is given by?

a) m = np
b) $m = (np)^2$
c) m = np(1-p)
d) m = p

Que. 17 If ‘m’ is the mean of a Poisson Distribution, then variance is given by ________

a) m^2
b) $m^{1/2}$
c) m
d) $m/2$

Que. 18 If ‘m’ is the mean of a Poisson Distribution, the standard deviation is given by _

a) $\sqrt{m}$
b) m^2
c) m
d) m/2

Que. 19 In a Poisson Distribution, the mean and variance are equal.
a) True
b) False

Que. 20 Poisson distribution is applied for ___________
a) Continuous Random Variable
b) Discrete Random Variable
c) Irregular Random Variable
d) Uncertain Random Variable

Que. 21 If 'm' is the mean of Poisson Distribution, the P(0) is given by___________
a) e^{-m}
b) e^{m}
c) e
d) m^{-e}

Que. 22 In a Poisson distribution, the mean and standard deviation are equal.
a) True
b) False

Que. 23 For a Poisson Distribution, if mean(m) = 1, then P(1) is?
a) 1/e
b) e
c) e/2
d) Indeterminate

Que. 24 The mean of hypergeometric distribution is _____________
a) n*k / N-1
b) n*k-1 / N
c) n-1*k / N
d) n*k / N

Que. 25 The Variance of hypergeometric distribution is given as __________
a) $n * k * (N – k) * (N – 1) / [N^2 * (N – 1)]$
b) $n * k * (N – k) * (N – n) / [N^2 * (N – k)]$
c) $n * k * (N – 1) * (N – n) / [N^2 * (N – 1)]$
d) $n * k * (N – k) * (N – n) / [N^2 * (N – 1)]$

Que. 26 Hypergeometric probability of hypergeometric distribution function is given by the formula _________
a) $h(x; N, n, k) = [{}^{k}C_{x}] [{}^{N}C_{n-x}] / [{}^{N}C_{n}]$
b) $h(x; N, n, k) = [{}^{k}C_{x}] [{}^{N-k}C_{n-x}] / [{}^{N}C_{n}]$
c) $h(x; N, n, k) = [{}^{k}C_{x}] [{}^{N-k}C_{n}] / [{}^{N}C_{n}]$
d) $h(x; N, n, k) = [{}^{k}C_{x}] [{}^{N-k}C_{n-x}] / [{}^{N-k}C_{n}]$

Que. 27 Normal Distribution is applied for ___________
a) Continuous Random Distribution
b) Discrete Random Variable
c) Irregular Random Variable
d) Uncertain Random Variable

Que. 28 The shape of the Normal Curve is ____________

a) Bell Shaped (b) Flat (c) Circular (d) Spiked

Que. 29 Normal Distribution is symmetric is about ____________

a) Variance

b) Mean

c) Standard deviation

d) Covariance

Que. 30 For a standard normal variate, the value of mean is?

a) ∞

b) 1

c) 0

d) not defined

Que. 31 The area under a standard normal curve is?

a) 0

b) 1

c) ∞

d) not defined

Que. 32 The standard normal curve is symmetric about the

valuea) 0.5

b) 1

c) ∞

d) 0

Que. 33 For a standard normal variate, the value of Standard Deviation is ____________

a) 0

b) 1

c) ∞

d) not defined

Que. 34 Normal Distribution is also known as____________

a) Cauchy's Distribution

b) Laplacian Distribution

c) Gaussian Distribution

d) Lagrangian Distribution

Que. 35 In Normal distribution, the highest value of ordinate occurs at ____________

a) Mean

b) Variance

c) Extremes

d) Same value occurs at all points

Que. 36 In Standard normal distribution, the value of mode is ____________

a) 2

b) 1

c) 0

d) Not fixed

Que. 37 In Standard normal distribution, the value of median is ____________

a) 1

b) 0

c) 2

d) Not fixed

- **Answer the following Questions in one Sentence each**

Que. 1 What is Discrete Random Variable?

Que. 2 What is Continuous Random Variable?

Que. 3 Give name of different discrete probability distribution studied in this chapter.

Que. 4 Give name of different Continuous probability distribution studied in this chapter.

Que. 5 State parameter of Binomial Probability Distribution.

Que. 6 If mean and variance of Binomial Distribution is 4 and 2.4 respectively, then what is probability of success?

Que. 7 What is the mean of Binomial Probability distribution?

Que. 8 Write a formula of Variance and Standard Deviation of Binomial Probability Distribution.

Que. 9 When we use Poisson Probability Distribution?

Que. 10 Write parameters of Poisson Probability Distribution.

Que. 11 If mean of Poisson distribution is 0.81 then what is the standard deviation of

Poisson Probability Distribution?

Que. 12 Write probability mass function of Poisson Probability Distribution.

Que. 13 Write formula of Variance of Poisson Probability Distribution.

Que. 14 Poisson probability distribution is used for discrete random variable or continuous random variable?

Que. 15 Write mean of Hyper Geometric Probability Distribution.

Que. 16 Is Hyper Geometric Distribution is used for continuous random variable?

Que. 17 State parameters of Hyper Geometric Distribution.

Que. 18 Write probability density function of Uniform probability distribution.

Que. 19 Mean of Uniform Distribution is________.

Que. 20 Normal Probability distribution is used for________________random variable.

Que. 21 Area under the Normal curve =_________.

Que. 22 Shape of the normal curve is_____________.

Que. 23 Give formula of mean deviation of Normal Probability Distribution.

Que. 24 Find the Quartile Deviation of Standard Normal Distribution.

- ## Answer the following Questions in detail

Que. 1 Write properties of Binomial Probability Distribution.

Que. 2 State uses of Binomial Probability distribution.

Que. 3 An unbiased coin is tossed for 6 time. Find the probabilities of getting (i) 5 heads (ii) at the most 2 heads.

Que. 4 The probability of occurrence of an occupational disease to a worker of a chemical factory is ¼. Find the probability that 2 out of 5 workers chosen at random will suffer from this disease.

Que. 5 The mean and variance of Binomial Probability Distribution are 15 and 6 respectively. Find the parameters of Binomial Probability Distribution.

Que. 6 There are two defective pencils in a pack of dozen pencils. If three pencils are taken at random, find the probabilities that (i) at the most one pencil is defective (ii) two pencils are defective.

Que. 7 For a Binomial variate $n = 10$ and $(x = 5) = 2. P(x = 4)$ find the value of p,

mean and variance.

Que. 8 The probability that a person hits a target is 1. Find the probability that he will hit

3

the target in 3 times out of 5 trials.

Que. 9 If 9 ships out of 10 ships safely reach the destination, find mean and standard

deviation of number of ships reaching safely to the destination out of 400 ships.

Que. 10 A six faced die is so constructed that the probability of getting an even number is twice the probability of getting an odd number. Find the probabilities that out of 5 trials (i) all the five will give even numbers (ii) 2 or 3 trials will give even

numbers.

Que. 11 A particular train reaches the destination in time in 75 pereent of the times. A person travels 5 times in that train. Find the probability that he will reach the

destination in time, for all the 5 times.

Que. 12 In a competitive test there are 6 objective questions, and three alternatives are given for each question. Only one of them is correct. A candidate does not know the correct answer of any of the questions and hence in each question he ticks any one of the alternatives randomly. Find the probabilities of getting (i) all correct

answers (ii) at least 4 correct answers.

Que. 13 Write properties of Poisson Probability Distribution.

Que. 14 State Application of Poisson Probability distribution.

Que. 15 There are 100 misprints in a book of 100 pages. If page is selected at random, find the probabilities that (i) there will be no misprint in the page (ii) there will be 1

misprint (iii) there will be at the most 2 misprints.

Que. 16 On an average 1.5 percent of electric bulbs are found to be defective in a bulb manufacturing factory. Using Poisson distribution find the probability of 4

defective bulbs in a box of 200 bulbs.

Que. 17 For a Poisson variate P(1) = P(2), find the parameter of Poisson Probability

dsitribution and P(0).

Que. 18 The probability that a blade manufactured by a factory is defective is 500. Blades are packed in packets of 10 blades. Find the probability of packets containing (i) no defective blade (ii) one defective blade (iii) 2 defective blades.

Que. 19 The probability that a patient will get the reaction to a particular injection is

0.001. 2000 patients are given that injection. Find the probabilities that (i) 3 patients will get a reaction (ii) more than 2 will get a reaction.

Que. 20 Between the hours of 2 and 4 p.m. the average number of phone calls per minute coming into the switch board of a company is 2.5. Find the probabilities that during one particular minute there will be (i) no phone call at all (ii) exactly 3

calls.

Que. 21 For a Poisson variate $3(x = 2) = (x = 4)$. Find mean and variance.

Que. 22 The standard deviation of Poisson variable is 0.8. find its mean, P(0) and P(1).

Que. 23 In the manufacturing of cotter pins it is known that 3% of the pins are defective. The pins are sold in boxes of 100 and it is guaranteed that not more than 4 pins will be defective in a box. What is the probability that a box will meet this

guarantee?

Que. 24 If 3% of electric bulbs manufactured by a company are defective, find the probability that in a sample of 100 bulbs, exactly 5 bulbs are defective.

Que. 25 In one hospital 3 percent of the patients demand special rooms. On a particular

day 3 special rooms were vacant. If 50 patients were admitted in the hospital on

that day, find the probabilities that (1) no patient demanded special room (ii) the demands for special room were not met.

Que. 26 In a book on an average there are 3 misprints in 5 pages. Using Poisson

distribution, find the number of pages having more than 2 misprints in that book of 100 pages.

Que. 27 A factory produces 0.5% defective articles. If a sample of 100 articles is taken

from the production, find the probability of getting 2 or more defective articles.

Que. 28 Write properties of Hyper Geometric Distribution.

Que. 29 A deck of cards contains 20 cards: 6 red cards and 14 black cards. 5 cards are drawn randomly without replacement. What is the probability that exactly 4 red

cards are drawn?

Que. 30 A small voting district has 101 female voters and 95 male voters. A random sample of 10 voters is drawn. What is the probability exactly 7 of the voters will be female?

Que. 31 Write properties of Uniform Distribution.

Que. 32 Write properties of Normal Probability Distribution.

Que. 33 State importance of Normal Probability Distribution.

Que. 34 The mean and standard deviation of marks of 500 students in an examination are 52 and 8 respectively. If the marks are normally distributed, find the probability that (i) students getting marks more than 60. (ii) students getting marks between 48 and 56. (iii) If the standard of passing is of 36 marks, find the probability of students failing in the examination.

Que. 35 In a normal distribution mean 21.5 and standard deviation 2.5. Find the following values: (i) $P\{18 \le x \le 25\}$ (ii) $P\{22 \le x \le 28\}$ (iii) $P\{x \ge 28\}$ (iv) $P\{x \le 18\}$

Que. 36 The average mark of 400 students in Statistics is 52 and s.d. of the marks is 8. If (i) the standard of passing is of 40 marks, (ii) the students securing marks between 48 and 60 are given second class, (iii) ateast 66 marks are necessary for getting distinction. Find the probability of students failing in the examination, getting second class and getting distinction.

Que. 37 The average weight of 1000 boys of a college is 52 kg. and its. s.d. is 3 kg. Assuming the weight to be normally distributed, find the probability of boys with weights (i) between 48 and 53 kg. (ii) exactly 56 kg.

Que. 38 The mean and standard deviation of a normal distribution are 30 and 5 respectively find the probabilities: (i) $26 \le x \le 40$ (ii) $x \ge 45$ (iii) $25 \le x \le 35$

Que. 39 The mean and standard deviation of a normal distribution are 20.5 and 5 respectively find median, mode, quartile deviation and mean deviation.

`

UNIT 10 PROBABILITY DISTRIBUTION OF RANDOM VARIABLE

10.1 INTRODUCTION

In studying a probability experiment, it is often useful to work with quantitative values to represent outcomes. These quantitative values associated to outcomes are called random variables. In this section, we explore random variables that take on numeric values that can be listed. For example, number of books, number of phone calls, number of student, etc. are discrete random variables. On the other hand, hair color, honesty, beauties are not a random variable because all those are not measure in numeric. Also, any decimal between 0 and 1 is not discrete because we cannot list out all the decimals. In analyzing real life data, we will apply fundamental concepts about discrete probability distributions to estimate likelihoods and draw inferences.

Probability is used to denote the happening of a certain event and the occurrence of that event, based on past experiences. The mathematical expectation is the events which are either impossible or a certain event in the experiment. Probability of an impossible event is zero, which is possible only if the numerator is 0. Probability of a certain event is 1 which is possible only if the numerator and denominator are equal.

10.2 MEAN FOR DISCRETE DISTRIBUTION – RANDOM VARIABLE (MATHEMATICAL EXPECTATION / EXPECTED VALUE)

Mean of discrete distribution is known as Mathematical expectation and also known as the expected value**,** which is the summation of all possible values from a random variable. It is also known as the product of the probability of an event occurring, denoted by P(x), and the value corresponding with the actually observed occurrence of the event.

For a random variable expected value is a useful property. E(X) is the expected value and can be computed by the summation of the overall distinct values that is the random variable. The mathematical expectation is denoted by the formula:

$$\mu = (x) = \sum\{x_i\ P(x_i)\}; \qquad i = 1,2,3,\dots,n$$

where, x is a discrete random variable with the probability P(x), where P(x) is probability of the occurrence of discrete random variable x and n is the number of all possible values.

The mean of discrete random variable (mathematical expectation) of an indicator variable can be 0 if there is no occurrence of an event A, and the mathematical expectation of an indicator variable can be 1 if there is an occurrence of an event A.

10.2.1 Properties of Expectation

1. If X and Y are the two variables, then the mathematical expectation of the sum of the two variables is equal to the sum of the mathematical expectation of X and the mathematical expectation of Y.

$$E(X+Y) = E(X) + E(Y)$$

2. The mathematical expectation of the product of the two random variables will be the product of the mathematical expectation of those two variables, but the condition is that the two variables are independent in nature. In other words, the mathematical expectation of the product of the n number of independent random variables is equal to the product of the mathematical expectation of the n independent random variables.

$$E(XY) = E(X) \times E(Y)$$

3. The mathematical expectation of the sum of a constant and the function of a random variable is equal to the sum of the constant and the mathematical expectation of the function of that random variable.

$$E\{a + f(X)\} = a + E\{f(X)\},$$

   where, a is a constant and f(X) is the function of discrete random variable x.

4. The mathematical expectation of the sum of product between a constant and function of a random variable and the other constant is equal to the sum of the product of the constant and the mathematical expectation of the function of that random variable and the other constant.

$$E\ (aX + b) = aE(X) + b,$$

   where, a and b are constants.

5. The mathematical expectation of a linear combination of the random variables and constant is equal to the sum of the product of ‘n’ constant and the mathematical expectation of the ‘n’ number of variables.

$$E(\sum aiXi)= \sum \{ai \times E(Xi)\}$$

Where, ai, (i=1…n) are constants.

6. The mathematical expectation of deviation taken from mean is always zero.

$$E(x - \mu) = E(x) - E(\mu) = \mu - \mu = 0$$

10.3 VARIANCE OF DISCRETE RANDOM VARIABLE

The average of the squares of the deviation from mean of a random variable x is said to be the variance of the discrete random variable x.

If the mean of a discrete random variable x is E(x) = μ then the variance of a discrete random variable x can be given as follows:

$$(x) = (x - \mu)^2 = E(x^2 - 2x\mu + \mu^2)$$

$$= (x^2) - 2(x) + (\mu^2) = (x^2) - 2\mu.\mu + \mu^2$$

$$= (x^2) - 2\mu^2 + \mu^2 = (x^2) - \mu^2$$

$$\boldsymbol{V(x) = E(x^2) - [E(x)]^2}$$

Some properties of variance of discrete random variable x:

(1) If k is constant V(k) = 0
(2) $(kx) = k^2(x)$
(3) $(ax + b) = a^2(x)$
(4) If x and y are independent variables $(ax + by) = a^2(x) + b^2V(y)$

10.4 REAL WORLD APPLICATIONS OF MATHEMATICAL EXPECTATION OF RANDOM VARIABLE X

Illustration 1: Find the mean of the number found on the dice when an unbiased dice is thrown.

Answer:

When we throw an unbiased dice we will get {1, 2, 3, 4, 5, 6} number on it. Probability of getting any number between 1 to 6 is equal and it is 1/6. To find mean we construct the following table.

Number on Dice (x)	Probability of Number on Dice (x)	x. P(x)
1	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{2}{6}$
3	$\frac{1}{6}$	$\frac{3}{6}$
4	$\frac{1}{6}$	$\frac{4}{6}$
5	$\frac{1}{6}$	$\frac{5}{6}$
6	$\frac{1}{6}$	$\frac{6}{6}$
	Total	$\frac{21}{6}$

$$\text{Mean } (\mu) = (x) = \sum\{x_i P(x_i)\} = \frac{21}{6} = 3.5$$

It is important to know that "mean value (expected value)" is not the same as "most probable value" and it is not necessary that it will be one of the probable values.

Illustration 2: Two dices are thrown. Find the mean (expected) value of the total on the dice.
Answer:

As we have thrown two dice we will get minimum total 2 and maximum total 12. So we may construct the following table to calculate mean value of the total on the dice.

Total on Dice (x)	Probability P(x)	x. P(x)
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36
	Total	252/36

`

Mean value on the two dice,

$$\mu = E(x) = \sum x.P(x) = \frac{252}{36} = 7$$

So, when two dice are thrown, then the mean (expected) value of the total on the dice is 7.

Illustration 3: 4 coins are tossed simultaneously; find the mean (expected) number of heads and its variance.

Answer:

As per the experiment we can derive all possible outcomes as,

HHHH, THHH, HTHT, TTHT,

HHHT, HHTT, THHT, THTT,

HHTH, TTHH, HTTH, HTTT,

HTHH, THTH, TTTH, TTTT

Here, random variable x = no. of heads

To find mean and variance of random variable we construct the following table.

No. of heads (x)	Probability P(x)	x. P(x)	x^2. P(x)
0	1/16	0	0
1	4/16	4/16	4/16
2	6/16	12/16	24/16
3	4/16	12/16	36/16
4	1/36	4/16	16/16
	Total	$\frac{32}{16} = 2$	$\frac{80}{16} = 5$

Mean of the random variable (x),

$$\mu = E(x) = \sum x.(x) = \frac{32}{16} = 2$$

Variance of the random variable (x),

$$(x) = (x^2) - [(x)]^2 = 5 - (2)^2 = 5 - 4$$

$$(x) = 1$$

Hence the mean and variance of random variable are 2 and 1 respectively.

Illustration 4: the probability distribution of a random variable x is as follows.

x_i	2	3	4	5	6	7	8	9	10
$P(x_i)$	**0.05**	**0.10**	**0.30**	**0.20**	**0.05**	**0.10**	**0.05**	**0.10**	**0.05**

Find mean and S.D. of x.

Note: S.D. (Standard deviation) is obtained by taking square root of variance.

Answer:

To compute mean and variance of random variable x, we construct the following table.

x_i	2	3	4	5	6	7	8	9	10	Total
$P(x_i)$	0.05	0.10	0.30	0.20	0.05	0.10	0.05	0.10	0.05	1
$x_i\, P(x_i)$	0.10	0.30	1.20	1.00	0.30	0.70	0.40	0.90	0.50	5.4
$x_i^2\, P(x_i)$	0.20	0.90	4.80	5.00	1.80	4.90	3.20	8.10	5.00	33.9

Mean of the random variable (x),

$$\mu = E(x) = \sum x.(x) = 5.4$$

Variance of the random variable (x),

$$(x) = (x^2) - [(x)]^2 = 33.9 - (5.4)^2 = 33.9 - 29.16$$

$$(x) = 4.74$$

Standard deviation of random variable (x),

$$SD = \sqrt{(x)} = \sqrt{4.74} \cong 2.18$$

Hence the mean and SD of x are 5.4 and 2.18 respectively.

Illustration 5: The probability distribution of a random variable is as follows:

x_i	15	16	17	18	19	20
$P(x_i)$	**0.04**	**0.19**	**3p**	**0.26**	**p**	**0.07**

Find the value of p and hence obtain expected value of x.

Answer:

As sum of all probabilities is 1.

`

$$\sum P(x_i) = 1$$

$$0.04 + 0.19 + 3p + 0.26 + p + 0.07 = 1$$

$$0.56 + 4p = 1$$

$$4p = 1 - 0.56$$

$$4p = 0.44$$

$$p = \frac{0.44}{4} = 0.11$$

Further,

x_i	15	16	17	18	19	20	Total
$P(x_i)$	0.04	0.19	0.33	0.26	0.11	0.07	1
$x_iP(x_i)$	0.60	3.04	5.61	4.68	2.09	1.40	17.42

Expected value of random variable x,

$$E(x) = \sum x.(x) = 17.42$$

Illustration 6: The probability distribution of a random variable is as follows:

x_i	-1	0	1	2	3	4
$P(x_i)$	**1/6**	**1/3**	**p**	**P**	**1/12**	**1/12**

Find the value of p and obtain the mean and variance of random variable x.

Answer:

$$\sum (x_i) = 1$$

$$\frac{1}{6} + \frac{1}{3} + p + p + \frac{1}{12} + \frac{1}{12} = 1$$

$$\frac{2+4+1+1}{12} + 2p = 1$$

$$\frac{8}{12} + 2p = 1$$

$$2p = 1 - \frac{8}{12} = 1 - \frac{2}{3} = \frac{1}{3}$$

$$p = \frac{1}{2(3)} = \frac{1}{6} = \frac{2}{12}$$

To make computation easy we did equal denominator to all probabilities i.e. for this illustration it is 12.

So,

x_i	-1	0	1	2	3	4	Total
$P(x_i)$	2/12	4/12	2/12	2/12	1/12	1/12	1
$\boldsymbol{x_i\, P(x_i)}$	-1/12	0	2/12	4/12	3/12	4/12	12/12 = 1
$\boldsymbol{x^2_i\, P(x_i)}$	1/12	0	2/12	8/12	9/12	16/12	36/12 = 3

Mean of the random variable (x),

$$\mu = E(x) = \sum x.(x) = 1$$

Variance of the random variable (x),

$$(x) = (x^2) - [(x)]^2 = 3 - (1)^2 = 3 - 1$$

$$(x) = 2$$

Hence, Mean and Variance of random variable x is 1 and 3 respectively.

Illustration 7: Two coins are tossed simultaneously, a person receives Rs. 8 for each head and losses Rs. 10 for each tail. Find the mean value of the amount gained by him.

Answer:

Here, in this experiment we have four different results mentioned as below.

Results	HH	HT	TH	TT
Amount (gained)	16	-2	-2	-20
*Negative sign indicate lose.				

To find the mean value of the amount gained by a person,

x_i	16	-2	-20	Total

`

$P(x_i)$	$\frac{1}{4} = 0.25$	$\frac{2}{4} = 0.50$	$\frac{1}{4} = 0.25$	1
$\boldsymbol{x_i\ P(x_i)}$	4	-1	-5	-2

Mean value of the amount gained by a person,

$$\mu = E(x) = \sum x.(x) = -2$$

So, on an average per experiment a person loses 2 rupees.

❖ **CHECK YOUR PROGRESS**

- **Answer the following Questions in one Sentence each**

Que. 1 Define discrete random variable.

Que. 2 What is the sum of probability?

Que. 3 Write a formula of mean of discrete random variable.

Que. 4 Write a formula to calculate variance of discrete random variable.

Que. 5 Write an addition property of expectation when two random variables are independent.

Que. 6 Write a multiplication property of expectation when two random variables are independent.

Que. 7 Write a formula to calculate standard deviation of discrete random variable.

Que. 8 Find the value of p from the following probability distribution.

x_i	12	13	14	15	16	17
$P(x_i)$	0.25	0.19	K	0.06	0.20	0.34

Que. 8 Define mathematical expectation.

Que. 9 Mean of discrete random variable is known as…………………..

- **Answer the following Questions in detail**

Que. 1 Define mean of random variable x in terms of discrete probability distribution.

Que. 2 Write properties of expected value of discrete random variable x.

Que. 3 Write properties of Variance of random variable x.

Que. 4 The probability distribution of a random variable x is as follows. Find,

(i) The value of p (ii) mean of random variable x.

x_i	0	1	2	3	4
$P(x_i)$	1/16	p	3/8	p	1/16

Que. 5 The probability distribution of a random variable x is as follows. Find (i) the value of k (ii) the probability distribution of x (iii) mean value of random variable x (iv) variance of random variable x.

x_i	0	1	2	3	4	5	6	7
$P(x_i)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2 + k$

Que. 6 The probability distribution of a random variable x is as follows.

x_i	-1	0	1	2	3	4
$P(x_i)$	1/8	1/8	1/4	1/4	1/8	1/8

Find the mean and variance of x.

Que. 7 The probability distribution of a random variable x is as follows.

x_i	0	1	2	3	4
$P(x_i)$	0.1	P	0.3	p	0.1

Find the value of p, mean and variance of x.

Que. 8 The probability distribution of demand of a commodity is given below:

Demand (x_i)	5	6	7	8	9	10
$P(x_i)$	0.05	0.1	0.3	0.4	0.1	0.05

Find the expected demand and its variance.

Que. 9 There are 6 slips in a box and numbers 1, 1, 2, 2, 3, 3 are written on these slips. Two slips are taken at random from the box, find the expected value of the sum of the numbers on thetwo slips.

Que. 10	There are two coins. On one face of each coin 1 is written and on the other face 2 is written. The coins are tossed simultaneously. Find the expected value of the total on the coins.
Que. 11	10,000 tickets each of Re. 1 are sold in a lottery. There is only one ticket in the lottery bearing a prize of Rs. 8000. A person is having one ticket of the lottery. Find hisexpectation.
Que. 12	There are 10 electric bulbs in a box in which 3 are defective bulbs. If 3 bulbs are selected at random from the box, find the expected number of defective bulbs.
Que. 13	There are 5 white and 3 black balls in a box. 3 balls are taken at random from the box. Find the expected number of black balls.

`

Que. 14	There are 4 black and 2 white balls in a box and 2 balls are taken at random from it. If aperson receives Rs. 4 for each white ball and loses Rs. 2 for each black ball, find the mathematical expectation of the amount received by him.
Que. 15	There are 3 black and 2 white balls in a box. 2 balls are taken from it. Rs. 24 is given for each black ball. What amount should be charged for each white ball so that the game isfair?
Que. 16	There are 8 screws in a packet of which 2 are defective. If 2 screws are taken at random,find the expected number of defective screws and also obtain its variance.

Que. 17 A random variable x has the following probability distribution,

x_i	3	4	5	6
$P(x_i)$	0.2	0.4	0.3	0.1

Find the mean and variance of a random variable x.

BAOU
Education for All

स्वाध्यायः परमं तपः

યુનિવર્સિટી ગીત

સ્વાધ્યાયઃ પરમં તપઃ
સ્વાધ્યાયઃ પરમં તપઃ
સ્વાધ્યાયઃ પરમં તપઃ

શિક્ષણ, સંસ્કૃતિ, સદ્ભાવ, દિવ્યબોધનું ધામ
ડૉ. બાબાસાહેબ આંબેડકર ઓપન યુનિવર્સિટી નામ;
સૌને સૌની પાંખ મળે, ને સૌને સૌનું આભ,
દશે દિશામાં સ્મિત વહે હો દશે દિશે શુભ-લાભ.

અભણ રહી અજ્ઞાનના શાને, અંધકારને પીવો ?
કહે બુદ્ધ આંબેડકર કહે, તું થા તારો દીવો;
શારદીય અજવાળા પહોંચ્યાં ગુર્જર ગામે ગામ
ધ્રુવ તારકની જેમ ઝળહળે એકલવ્યની શાન.

સરસ્વતીના મયૂર તમારે ફળિયે આવી ગહેકે
અંધકારને હડસેલીને ઉજાસના ફૂલ મહેંકે;
બંધન નહીં કો સ્થાન સમયના જવું ન ઘરથી દૂર
ઘર આવી મા હરે શારદા દૈન્ય તિમિરના પૂર.

સંસ્કારોની સુગંધ મહેંકે, મન મંદિરને ધામે
સુખની ટપાલ પહોંચે સૌને પોતાને સરનામે;
સમાજ કેરે દરિયે હાંકી શિક્ષણ કેરું વહાણ,
આવો કરીયે આપણ સૌ
ભવ્ય રાષ્ટ્ર નિર્માણ...
દિવ્ય રાષ્ટ્ર નિર્માણ...
ભવ્ય રાષ્ટ્ર નિર્માણ

DR. BABASAHEB AMBEDKAR OPEN UNIVERSITY
(Established by Government of Gujarat)
'Jyotirmay' Parisar,
Sarkhej-Gandhinagar Highway, Chhârodi, Ahmedabad-382 481
Website : www.baou.edu.in