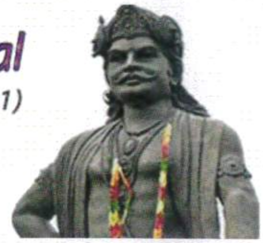
Madhya Pradesh Bhoj (Open) University, Bhopal

(Established under an Act of State Assembly in 1991)

मध्यप्रदेश भोज (मुक्त) विश्वविद्यालय, भोपाल



SELF - LEARNING MATERIAL

MBA, Second Year

Paper – II

QUANTITATIVE TECHNIQUES FOR MANAGERS

MBA Second Year

Paper - II

QUANTITATIVE TECHNIQUES FOR MANAGERS

मध्यप्रदेश भोज (मुक्त) विश्वविद्यालय – भोपाल

MADHYA PRADESH BHOJ (OPEN) UNIVERSITY - BHOPAL

COURSE WRITERS

Dr Chandan Maheshkar, Senior Consultant, Centre for Internal Quality Assurance (CIQA), Madhya Pradesh Bhoj (Open) University, Bhopal, and Former Research Fellow, Management, Directorate of Distance Education (DDE), University of Indore

Dr JS Chandan, Prof. Medgar Evers College, City University of New York, New York

CR Kothari, Ex- Associate Professor, Department of Economic Administration & Financial Management, University of Rajasthan, Rajasthan

S Kalavathy, Dean, Department of Science and Humanities, K Ramakrishnan College of Technology, Samayapuram, Trichy

VIKAS®

Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: A-27, 2nd Floor, Mohan Co-operative Industrial Estate, New Delhi 1100 44
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

SYLLABI-BOOK MAPPING TABLE

Quantitative Techniques for Managers

CONTENTS

INTRODUCTION

NOTES

Managers in all types of business organizations — whether in the private or public sector, manufacturing or service industry — have to take decisions on how best the organization's resources can be allocated. Such decisions are made on the basis of information obtained using statistical and decision-making techniques called quantitative methods. The analysis can be quantitative or numerical in nature, and today's managers should be able to assess, analyse and utilize these analytical tools. This book aims to explain some methods and techniques which are applied to quantitative information.

Quantitative techniques are powerful tools through which production can be augmented, profits maximized, costs minimized and production methods oriented for accomplishing certain predetermined objectives. Nowadays, there is a growing tendency to turn to quantitative methods as a means for solving various problems that arise in a business or industrial enterprise. The success of these tools has attracted more and more students, business executives and public administrators to the study of these techniques.

The objectives of this book are to:

- Introduce certain mathematical and basic quantitative methods frequently used in business decision-making.
- Explain how to assess these techniques and use them when appropriate.
- Formulate a suitable quantitative problem, obtain a solution and interpret important features of this solution in a business decision-making process.

Thus, quantitative techniques are those statistical and operations research or programming methods which help in the decision-making process, especially concerning business and industry. These methods involve the use of quantities to supplement judgement and intuition. They evaluate planning factors and alternatives as and when they arise, rather than prescribe courses of action. As such, quantitative methods may be defined as the methods providing the decision-maker with a systematic and powerful means of analysis and help, based on mathematical and quantitative data, in exploring policies for achieving predetermined goals. These methods are particularly relevant to problems of complex business enterprises.

Quantitative techniques, have gained increasing importance since World War II in the technology of business administration. Management experts, while talking about quantitative techniques in modern times, usually have quantitative research methods in mind, which help in tackling intricate and complex problems of modern business and industry. Quantitative techniques for decision-making are, in fact, examples of the use of the scientific method of management. At times, there may be a problem of finding an acceptable definition of long-range company objectives. Management may be confronted with different viewpoints—some may stress the desirability of maximizing net profits, whereas others may focus attention primarily on the minimization of costs. Quantitative techniques can help resolve such dilemmas by permitting systematic evaluation of the best strategies for attaining

NOTES

different objectives. These techniques can also be used for estimating the worth of technical innovations as also of potential profits associated with the possible changes in rules and policies. Quantitative methodology serves to develop a scientific basis for coping with the uncertainties of future demand and for dealing with such problems where quantitative techniques can be used to generate a 'least risk' plan. Therefore, quantitative techniques render a valuable service in the field of business management. It is hoped that the wide coverage of the book in such an illustrative manner would be of great value to the students at large.

This book, *Quantitative Techniques For Managers* is divided into five units that follow the self-instruction mode with each unit beginning with an Introduction to the unit, followed by an outline of the Objectives. The detailed content is then presented in a simple but structured manner interspersed with Check Your Progress Questions to test the student's understanding of the topic. A Summary along with a list of Key Terms and a set of Self-Assessment Questions and Exercises is also provided at the end of each unit for recapitulation.

UNIT 1 DATA COLLECTION, PRESENTATION AND ANALYSIS

NOTES

Structure

1.0 INTRODUCTION

Every day we are confronted with some form of statistical information through different sources. All raw data cannot be termed as statistics. Similarly, single or isolated facts or figures cannot be called statistics as these cannot be compared or related to other figures within the same framework. Hence, any quantitative and numerical data can be identified as statistics when it possesses certain identifiable characteristics according to the norms of statistics. In every field of commerce, data collection methods play a valuable role in analysing the various elements of trade such as profit, loss, demand and supply of a commodity, resources available, recruiting people, etc. Accurate records have to be kept to keep the people updated about the current scenario of the society. As there are several methods of data collection, the methods that consume the least amount of time are put into use. Data collecting techniques such as questionnaires and interviews play a vital role in collecting large amount of information in a short period of time and hence have been discussed in this chapter. Experiments are resorted to when it is required to collect factual data when nothing is available for reference. It may also be conducted to verify a theory. Experiment is a study conducted under controlled conditions. In any type of business firm, large amount of raw data is generated from various business sources. Such amount of data becomes quite cumbersome and confusing for the management to handle and analyse. In a business firm, data can be of various types, relating to various categories such as number of each item of the inventory, record of sales from different departments, keeping an account of all kinds of bills and so on. It is almost impossible for the management to deal with all this data in raw form. Therefore, such data must be presented in a suitable and summarized form without any loss of relevant information so that it can be efficiently used for decision-making. Hence, we construct appropriate tables, graphs and diagrams to interpret and summarize the entire set of raw data.

Correlation analysis technique looks at the indirect relationships and establishes the variables which are most closely associated with a given data or mindset. It is the process of finding how accurately the line fits using the observations. Correlation analysis can be referred as the statistical tool used to describe the degree to which one variable is related to another. The relationship, if any, is usually assumed to be a linear one. In fact, the word correlation refers to the relationship or the interdependence between two variables. There are various phenomena which have relation to each other. The theory by means of which quantitative connections between two sets of phenomena are determined is called the 'Theory of Correlation'. On the basis of the theory of correlation, you can study the comparative changes occurring in two related phenomena and their cause-effect relation. Thus, correlation is concerned with relationship between two related and quantifiable variables and can be positive or negative.

Regression analysis technique is used to determine the statistical relationship between two or more variables and to make prediction of one variable on the basis of one or more other variables. While making use of the regression techniques for making predictions, it is always assumed that there is an actual relationship between the dependent and independent variables. The variable to be predicted is called the dependent variable and the variable on which the prediction is based is called the independent variable. You will also learn about the scatter diagram, least squares method and standard error of the estimate. Standard error of estimate is a measure developed by the statisticians for measuring the reliability of the estimating equation. The larger the standard error (SE) of estimate, the greater happens to be the dispersion, or scattering, of given observations around the regression line. But if the SE of estimate happens to be zero then the estimating equation is a 'perfect' estimator, i.e., cent per cent correct estimator of the dependent variable.

This unit introduces the basic principles and concepts of accounting. It further highlights the different systems of bookkeeping and accounting. In this unit, you will learn about various data collection methods. This unit describes the advantages and shortfalls of various types of observation. You will also learn about the process of preparing a questionnaire, what all should be kept in mind while drafting it and what pattern of questions should be adopted, i.e., dichotomous, multiple choice or open questions. Also, you would learn about the different modes of interviews along with their merits and demerits. In this unit, you will learn about the construction of tables, diagrams and graphs, how important these are to a business and what their usages are. In view of the ever increasing importance of statistical data in business operations and their management, this unit discusses the presentation of data in the form of graphs, tables and diagrams, their importance and use. In this unit, you will study arithmetic procedures that can be used for analysing and interpreting quantitative data. These measures and procedures relate to some properties and characteristics of data which include measures of central location of data, other measures of non-central location, measures of dispersion of data in itself and around the mean and the shape of the data. In this unit you will also understand the concept of arithmetic mean, median and mode. You will also learn to calculate geometric mean. The unit also explains quartiles, deciles and percentiles. You will also learn to explain data grouped into a frequency distribution. This unit discusses the measures of dispersion, which in itself is a very important property of a distribution and needs to be measured by appropriate statistics. Hence, this unit has taken into consideration several aspects of dispersion. It describes absolute and relative measures of dispersion. It deals with range, the crudest measure of dispersion. It also explains quartile deviation, mean deviation and standard deviation. The unit also discusses variance, the square of deviation. In this unit, you will learn about correlation analysis and regression analysis.

1.1 OBJECTIVES

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

- Analyse the limitations of statistics and its relation with business and industry
- Explain the different types of data collecting methods
- Discuss the advantages and limitations of direct personal observation
- State the characteristics and elements of a questionnaire
- Examine the merits and shortfalls of using a questionnaire
- Describe the significance of interviews
- Explain the types of tables, graphs and diagrams
- Describe the concept of frequency polygon and relative frequency
- Explain the construction of ogive curve and their types
- Describe the use of histograms
- Evaluate and represent data in diagrammatic and graphic forms
- Calculate the arithmetic mean, geometric mean, mode and median
- Evaluate and represent the data using quartiles, deciles and percentiles
- Describe variance and coefficient of variation
- Explain correlation analysis and regression analysis
- Evaluate coefficient of determination and coefficient of correlation
- Calculate and interpret standard error of estimate

1.2 STATISTICS: BASIC CONCEPTS

Statistics originated from two quite dissimilar fields, viz., games of chance and political states. These two different fields are also termed as two distinct disciplines—one primarily analytical and the other essentially descriptive. The former is associated with the concept of chance and probability and the latter is concerned with the collection of data.

The theoretical development of the subject has its origin in the mid-17 century and many mathematicians and gamblers of France, Germany and England are credited for its development. Notable amongst them are Pascal (1623–1662), who investigated the properties of the coefficients of binomial expansion and James

NOTES

Bernoulli (1654–1705). who wrote the first treatise on the theory of probability.

As regards the descriptive side of statistics it may be stated that statistics is as old as statecraft. Since time immemorial men must have been compiling information about wealth and manpower for purpose of peace and war. This activity considerably expanded at each upsurge of social and political development and received added impetus in periods of war.

The development of statistics can be divided into the following three stages:

The empirical stage (1600). During this, the primitive stage of the subject, numerical facts were utilized by the rulers, principally as an aid in the administration of Government. Information was gathered about the number of people and the amount of property held by them—the former serving the ruler as an index of human fighting strength and the latter as an indication of actual and potential taxes.

The comparative stage (1600–1800). During this period statisticians frequently made comparisons between nations with a view to judging their relative strength and prosperity. In some countries enquiries were instituted to judge the economic and social conditions of their people. Colbert introduced in France a 'mercantile' theory of government whose basis was essentially statistical in character. In 1719, Frederick William I began gathering information about population occupation, house-taxes, city finance, etc., which helped to study the condition of the people.

*The modern st*age (1800 up to date). During this period statistics is viewed as a way of handling numerical facts rather than a mere device of collecting numerical data. Besides, there has been a considerable extension of the field of its applicability. It has now become a useful tool and statistical methods of analysis are now being increasingly used in biology, psychology, education, economics and business.

Limitations of Statistics

That statistical technique, because of its flexibility and economy, is growing in popularity and is being successfully employed by the seekers of truth in numerous fields of learning is a fact that cannot be denied. But it is not without limitations. It cannot be applied to all kinds of phenomena and cannot be made to answer all our queries.

Statistics deals with only those subjects of inquiry which are capable of being quantitatively measured and numerically expressed. This is an essential condition for the application of statistical methods.

Now all subjects cannot be expressed in numbers. Health, poverty, intelligence (to name only a few) are instances of objects that defy the measuring rod, and hence are not suitable for statistical analysis. It is true that efforts are being made to accord statistical treatment to subjects of this nature also. Health of the people is judged by a study of its death rate, longevity of life and the prevalence of any disease or diseases. Similarly intelligence of the students may be compared on the basis of marks obtained by them in a class test. But these are only indirect methods of approaching the problem and subsidiary to quite a number of other considerations which cannot be statistically dealt with.

Statistics deals only with aggregates of facts and no importance is attached to individual items. It is, therefore, suited only to those problems where group characteristic are desired to be studied. But where the knowledge about individual cases is necessary statistical techniques prove inadequate. The per capita consumption of food grains in a state will camouflage cases of starvation, if any. The scarcity felt by the poorer section may be more than made up by the extravagance of the rich. In such cases, therefore, statistics, will fail to reveal the real position.

Statistical data is only approximately and not mathematically correct. Greater emphasis is being laid on the sampling technique of collecting data. This means that by observing, only a limited number of items we make an estimate of the characteristic of the entire population. This system works well so long as the mathematical accuracy is not essential. But when exactness is essential statistics will fail to do the job.

Statistics can be used to establish wrong conclusions and, therefore, can be used only by experts. Since many of the statistical conclusions are based on sample studies, it is very common to come to wrong conclusions if one is not very careful about the techniques of analysis. In fact, one is so often deceived by 'correct' facts that there is a general distrust of things 'proved statistically'. Usually, most of these can be traced to incorrect application of methods. The next few sections illustrate some of the common statistical fallacies.

Statistical Method

Statistical approach to a problem may broadly be summarized as: (i) collection of facts: (ii) organization of facts; (iii) analysis of facts; and (iv) interpretation of facts.

A detailed discussion of the various methods of collection, presentation, analysis and interpretation of facts is given later in the unit. Here the intention is to give only a bird's eye-view of the entire statistical procedure,

(i) Collection of facts is the first step in the statistical treatment of a problem. Numerical facts are the raw materials upon which the statistician is to work and just as in a manufacturing concern the quality of a finished product depends, *inter alia,* upon the quality of the raw material, in the same manner, the validity of statistical conclusions will be governed, among other considerations, by the quality of data used. Assembling of the facts is thus a very important process and no pains should be spared to see that the data collected are accurate, reliable and thorough. One thing that should be noted here is that the work of collecting facts should be undertaken in a planned manner. Without proper planning the facts collected may not be suitable for the purpose and a lot of time and money may be wasted.

(ii) The data so collected will more often than not be a huge mass of facts running into hundreds and thousands of figures. Human mind has its limitations. No One can appreciate at a glance or even after a careful study hold in mind the information contained in a hundred or a thousand schedules. For a proper understanding of the data their irregularities must be brushed off and their bulk be reduced, *i.e.,* some process of condensation must take

NOTES

place. Condensation implies the organization, classification, tabulation and presentation of the data in a suitable form.

(iii) The process of statistical analysis is a method of abstracting significant facts from the collected mass of numerical data. This process includes such things as 'measures of central tendency'—the determination, of Mean, Median and Mode—'measures of dispersion' and the determination of trends and tendencies, etc. This is more or less a mechanical process involving the use of elementary mathematics.

(iv) The interpretation of the various statistical constants obtained through a process of statistical analysis is the final phase or the finishing process of the statistical technique. It involves those methods by which judgments are formed and inferences obtained. To make estimates of the population parameters on the basis of sample statistics in an example of the problem of interpretation. For the interpretation of results a knowledge of advanced mathematics is essential.

Relationship with Business and Industry

The need for statistical information in the smooth functioning of an undertaking increases along with its size. The bigger the concern the greater the need for statistics. In the era preceding the Industrial Revolution the master craftsman was in intimate touch with the sources of the supply of raw materials. He worked in his own home with the help of the members of his family and a few other employees whom he knew rather well. His customers were few and he knew them all personally. Thus, he had almost all information about his business and obviously no technique for the supply of this information was necessary.

Today too in an era of mass production technology, the business executive needs all such information for the successful conduct of affairs. But he cannot, even if he were to try, get this information in the same manner as the master craftsman did. Naturally, therefore, he has to resort to the statistical technique and statistics takes the place of personal observation. 'For better or worse, the modern business executive is largely dependent on statistical data and methods of analysis for essential information'. No business, large or small, public or private, can flourish in these days of large-scale production and cut-throat competition without the help of statistics. Statistical information is needed from the time the business is launched tilt the time of its exit. At the time of the floatation of the concern facts are required for the purposes of drawing up the financial plan of the proposed unit. All the factors that are likely to affect judgment on these matters are quantitative weighed and statistically analyzed before taking any decisions. A shrewd manufacturer must know in advance 'how much is to be produced,' 'how many workers and how much raw material is needed to produce that estimated quantity and what quality, type size, colour or grade of the product is to be manufactured.' In short, he must have a production plan. Now such a plan cannot be framed without quantitative facts. Statistics thus help in planning and formulation of future policies. Quantitative data will have to be collected and analysed, if a workable personnel plan is to be carried out. The only route for a personnel officer or a labour officer to get acquainted with the labour force numbering hundreds, or thousands, or even lakes, is to know its members through statistical analysis of information, largely quantitative. Wage levels and wage standards also require the statistical study of different jobs within the same organisation and the study of wages in like business undertakings.

In a labour dispute it is the official of the union who generally represents the workers. It is through statistical data that a man representing the workers knows about the working conditions, rates of wages, frequency of lock-outs, monthly earnings and other matters in the industry where the dispute may arise. Again, in negotiation conferences, proper data, competently collected and honestly analysed, may lead to an early and just solution of the differences.

Statistical methods of analysis are helpful in the marketing function of an enterprise through its enormous help in market research, advertisement campaigns and in comparing the sales performances. Statistics also directs attention towards the effective use of advertising funds.

Above all, statistical methods of analysis provide an important tool to management for cost and budgetary control. The most elementary use to the management is in the balancing of the activities of one part of a system against those of another, to secure that supplies equal requirements and that there are no 'bottlenecks' or parts that are not employed to the full.

Various statistical techniques viz., index numbers and analysis of time series help in the study of price behaviour; correlation and regression help in the estimation of relationships between dependent and one or more independent variables, *e.g.,* relationships are established between market demand and per capita income, inputs and outputs, etc.

The theory and technique of sampling can be used in connection with various business surveys with a considerable saving in time and money. Likewise these techniques are now being extensively used in test checking of accounts.

Statistical quality control is now being used in industry for establishing quality standards for products, for maintaining the requisite quality, and for assuring that the individual lots sold are of a given standard of acceptance.

Statistics is thus a useful tool in the hands of the management. But it must be remembered that no volume of statistics can replace the knowledge and experience of the executives. Statistics supplements their knowledge with more precise facts than were hitherto available.

The following are some typical situations in business which can be analyzed using statistical techniques:

- **Survey of consumer tastes:** To predict the acceptability of a synthetic soft-drink concentrate, a manufacturer distributed code- marked samples along with the samples of a leading brand. By analysing consumer preferences, he could select the sales territory in which to concentrate his effort, and obtained very good results.
- **Quality control:** An electric lamp manufacturer wanted to control the average life of his product. Since he obviously could not test every bulb to burnout, he devised a sampling plan wherein 25 out of every

lot of 1,000 bulbs were tested. The average life of these 25 bulbs provided a check on the quality of the whole lot.

- **Optimum inventory size:** Large dealerships require one to maintain a stock large enough to service all customers but not so large that money is tied up unnecessarily in idle stock. An auto spare parts manufacturer solved this problem by collecting statistics over a year and determining the probable distribution of demand, and calculated the optimum inventory levels.
- **Overbooking:** Because of passengers dropping out at the last moment, all airlines overbook their flights in the hope that some passengers will not show up. This is a risky proposition because if all do show up, one has a lot of irate passengers on one's hand. By proper statistical analysis one can determine the optimum overbooking so as to minimize this risk and yet maximize the load factor.

Check Your Progress

1. Name the two domains from which statistics originated.
2. How are the statistical methods of analysis helpful in the marketing function of an enterprise?

1.3 DATA COLLECTION

Observation may be defined as recording behavioural patterns without verbal communication.

Primary data can be collected using the following method.

(i) Direct personal observation

Under this method, the investigator presents himself personally before the informant and obtains a first hand information. This method is most suitable when the field of enquiry is small and a greater degree of accuracy is required.

The following are the merits and limitations of the observation method:

Merits

(*i*) The first hand information obtained by the investigator is bound to be more reliable and accurate since the investigator can extract the correct information by removing doubts, if any, in the minds of the respondents regarding certain questions.

(*ii*) High response rate, since the answers to various questions are obtained on the spot.

(*iii*) It permits explanation of questions concerning difficult subject matters.

(*iv*) It permits evaluation of respondent, his circumstances and reliability.

(*v*) This method is useful where spontaneity of response is required.

(*vi*) It provides personal rapport that helps to overcome reluctance to respond.

(*vii*) Where the investigator and the informant talk face to face, it becomes possible to explore questions in depth.

(*viii*) Information is collected promptly and there is no dribbling in.

Limitations

(*i*) This method is suitable only for intensive studies and not for extensive enquiries.

(*ii*) This method is time-consuming and the investigation may have to be spanned over a long period.

(*iii*) This method is highly subjective in nature and the results of the enquiry may be adversely affected by the personal biases, whim and prejudices of the investigator.

(ii) Questionnaire

The questionnaire method can be used either by mail or through enumerators.

(a) Mailed questionnaire method

Under this method, the investigator prepares a questionnaire containing a number of questions pertaining to the field of enquiry. These questionnaires are sent by post to the informants together with a polite covering letter explaining in detail the aims and objectives of collecting the information and requesting the respondents to cooperate by furnishing the correct replies and returning the questionnaire duly filled in. In order to ensure quick response, the return postage expenses are usually borne by the investigator. This method is usually adopted by research workers, private individuals and non-official agencies. The success of this method depends upon the proper drafting of the questionnaire and the cooperation of the respondents.

Merits

(*i*) By this method, a large field of investigation may be covered at a very low cost. In fact, this is the most economical method in terms of time, money and manpower.

(*ii*) Errors due to personal bias of the investigators or enumerators are completely eliminated as the information is supplied by the person concerned in his own handwriting.

Limitations

(*i*) This method can be used only if the respondents are educated and can understand the questions well, and reply in their own handwriting.

(*ii*) Sometimes, the informants may not send back the questionnaires and even if they return the schedules, they may be incorrectly filled in.

(*iii*) Sometimes, the informants are not willing to give written information in their own handwriting on certain personal questions like income, personal habits and property.

(*iv*) There is no scope for asking supplementary questions for cross-checking of the information supplied by the respondents.

(b) Questionnaire sent through enumerators

Under this method, instead of sending the questionnaire through post, the investigator appoints agents known as enumerators, who go to the respondents personally with the questionnaire, ask them the questions given therein, and record their replies.

This method is generally used by business houses, large public enterprises and research institutions.

Merits

(*i*) The information collected through this method is more reliable as the enumerators can explain in detail the objectives and aims of the enquiry to the respondents and win their cooperation.

(*ii*) Since the enumerators personally call on the respondents, there is very little non-response.

(*iii*) This technique can be used with advantage even if the respondents are illiterate.

(*iv*) The enumerators can effectively check the accuracy of the information supplied through some intelligent cross-questioning by asking supplementary questions.

Limitations

(*i*) The method is more expensive and can only be used by financially strong bodies or institutions.

(*ii*) It is more time-consuming than the mailed questionnaire method.

(*iii*) The success of the method depends upon the skill and efficiency of the enumerators to collect the information and also on the efficiency and wisdom with which the questionnaire is drafted.

Drafting or framing the questionnaire

Since the questionnaire is the only medium of communication between the investigator and the respondents, it must be designed or drafted with utmost care and caution so that all the relevant and essential information for the enquiry may be collected without any difficulty, ambiguity or vagueness. Designing of questionnaire, therefore, requires a high degree of skill and experience on the part of the investigator. No hard and fast rules can be laid down for designing or framing a questionnaire. However, much useful purpose would be served if the following general points are borne in mind while drafting a questionnaire:

- The size of the questionnaire should be as small as possible. The number of questions should be kept to the minimum keeping in view the nature, objectives and purpose of enquiry. Respondents' time should not be wasted by asking irrelevant and unimportant questions. Fifteen to twenty- five may be regarded as a fair number. If a larger number of questions is unavoidable in any enquiry, the questionnaire should preferably be divided into two or more parts.
- Questions should be clear, brief, unambiguous, non-offending, courteous in tone, corroborative in nature and to the point.
- Questions should be logically arranged.
- Questions should be short, simple and easy to understand. The usage of vague or multiple meaning words should be avoided. Unless the respondents are technically trained, the use of technical terms should be avoided.

NOTES

- Questions should be so designed that the respondents can easily comprehend and answer them. Questions involving mathematical calculations should not be asked.
- Questions of sensitive or personal nature should be avoided.
- The questionnaire should provide necessary instructions to the enumerators.
- If a particular question needs clarification, it should be explained by way of a footnote.
- Questions should be capable of objective answer. Various types of questions in the questionnaire may be grouped under three categories:
    - o ***Dichotomous or simple alternate questions*** in which the respondent has to choose between two clear-cut alternatives like 'Yes' or 'No', 'Right' or 'Wrong', 'Either', 'Or', and so on. This technique can be applied elegantly in situations where two clear-cut alternatives exist.
    - o ***Multiple choice questions*** in which the respondent is asked to select one out of a number of responses. All possible answers to a question are listed and the respondent chooses one of these. Such questions save time and facilitate tabulation. This method should be used only if a few alternative answers exist to a particular question.
    - o ***Open questions*** are those in which no alternative answers are suggested and the respondents are free to express their frank and independent opinions on the problem in their own words usually in essay form.
- **Cross-checks:** The questionnaire should be so designed as to provide a cross-check on the accuracy of the information supplied by the respondents by including some connected questions.
- **Pre-testing the questionnaire:** The questionnaire should be tried on a small group before using it for the given enquiry. This will help in improving or modifying the questionnaire in the light of drawbacks, shortcomings and problems faced by the investigator in the pre-test.
- A covering letter, stating briefly the aims and objects of the enquiry, soliciting cooperation of the respondents, and explaining various terms and concepts, should be enclosed along with the questionnaire.
- In case of mailed questionnaire method, a self-addressed stamped envelope should be enclosed.
- To ensure quick response, the respondents may be offerred incentives in the form of gift coupons, a sample of the product to be introduced, or a promise to supply a copy of the findings after the survey work is over.
- Method of tabulation and analysis, whether hand-operated, machine-operated or computerized, should also be kept in mind while designing the questionnaire.
- Lastly, the questionnaire should be made attractive by a proper layout and an appealing get up.

NOTES

A specimen questionnaire

This hypothetical study is adapted from a study developed by Deepak Mahendru in India. Assume that this study involves 200 professors in New York colleges who are asked about their interest in buying automobiles. The basic objective of this survey is to determine certain marketing trends among the population of professors in New York regarding their automobile buying patterns and are based upon the following factors:

- The profile of the decision maker who finally decides to buy a particular type of car.
- People around the decision maker who influence the decision-making process.
- The factors affecting the selection of a particular dealer of cars.
- People in the family who make or affect decisions regarding the maximum budget that can be allocated for purchasing a car.
- The effect of various options available in the car.
- The image and reliability of the company that makes these cars.
- The effect of heavy promotion on television about the utility of the car on the decision-maker.

(For the sake of simplicity, it is assumed that the professors have only one car in the family.)

The Questionnaire

1. General

    Name: ....

    Age: ....

    Sex: *M* .... *F* ....

    Marital status: Married .... Unmarried ....

    Number of members in the family

    1–2..................

    3–4..................

    5–6..................

    Over 6.............

    Yearly income

    Less than $30,000..................

    $30,000–$39,999.....................

    $40,000–$49,999.....................

    $50,000 and more..................

2. What type of car do you own now?

    ....American

    ....Japanese

NOTES

.................European

3. What size of car do you own?

   ....Luxury

   ....Mid-size

   ....Compact

4. Did you buy this car new or used?

   ....New....................Used

5. If you bought a used car, did you buy it from a dealer or a private party?
   ....Dealer.................Private party

6. If you bought a new car, how long have you owned this car?

   ....Number of years

7. If you bought a used car, how old is this car now?

   ....Number of years

8. Price paid for the car..........New..........Used

9. Who influenced your decision to purchase the above brand of car? Indicate if more than one.

   ....Yourself .... Your wife

   ....Your children .... Your friend

   ....Your neighbour .... Your colleague

   Others.. .

10. Indicate as to who decided about the budget allocation for the car.

    ....Yourself

    ....Your spouse

    ....Family decision

11. If you bought your car from a dealer, then who influenced your decision regarding the selection of a particular dealer?

    ....Yourself

    ....Your friend

    ....Your colleague

    ....Family decision

12. How did you come to know about this dealer?

    ....TV commercial

    ....Newspapers

    ....Personal references

    ....Others

13. Rank the following factors that affected the final decision at the time of purchasing the car. A rank of 1 measures the most important factor, a rank of 2 measures the second most important factor, and so on.

 ....Very inconvenient without the car

 ....Money was available

 ....Reputation of car manufacturer

 ....Discounts offered

 ....Interest rate on financing

 ....Guarantees and warranties offered

 ....Others

14. Did you make an extensive survey regarding price comparisons after you decided to buy the particular car? .... Yes......... No

15. If you bought a used car, how did you learn about it?............ Newspapers

 ....Friend .... Others

16. In order of preference, what were the major reasons for buying a used car?

 ....Unavailability of adequate funds

 ....Cheaper insurance

 ....Lack of parking garage

 ....Condition of the car

 ....Others

17. Which of the following media you think is most effective in creating an impact on the potential customer relative to a particular brand of the car?

 ....TV ....Newspapers

 ....Magazines ....Favourable news reports

 ....Word of mouth ....Others

The responses to such questions would form the basis of analysis in order to achieve the set marketing objectives.

(iii) Interviews

These constitute the following:

(a) Indirect personal interview

Under this method, instead of directly approaching the informants, the investigator interviews several third persons who are directly or indirectly concerned with the subject matter of the enquiry and who are in possession of the requisite information. Such a procedure is followed by the enquiry committees and commissions appointed by the Government of India. The committee selects persons, known as witnesses and collects information from them by getting answers to questions decided in advance. This method is highly suitable where direct personal investigation is not practicable either because the informants are unwilling or reluctant to supply information or where the information desired is complex and the study in hand is extensive.

Merits

- This method is less costly and less time-consuming than direct personal investigation.
- Under this method, the enquiry can be formulated and conducted more effectively and efficiently as it is possible to obtain the views and suggestions of the experts on the given problem.

Limitations

The success of this method depends upon the following:

- The representative character of the witnesses.
- The personal knowledge of the witnesses about the subject matter of enquiry.
- The personal prejudices of the witnesses as regards definiteness in stating what is wanted.
- The ability of the interviewer to extract information from the witnesses by asking appropriate questions and cross-questions.

(iv) Other methods

(a) Telephone survey. Under this method, the investigator, instead of presenting himself before the informants, contacts them on telephone and collects information from them.

Merits

- The method is more convenient than personal interview.
- This method is less time-consuming and can be applied even to extensive fields of enquiries. Telephone survey has all the other merits of personal interview.

Limitations

- This method excludes those who do not have a telephone and also those who have unlisted telephones.
- This method is also subjective in nature and personal bias, whim and prejudices of the investigator may adversely affect the results of the enquiry.

(b) Information received through local agents

Under this method, the information is not collected formally by the investigator, but local agents, commonly known as correspondents are appointed in different parts of the area under investigation. These agents collect information in their areas and transmit the same to the investigator. They apply their own judgement as to the best method of obtaining information. This method is usually employed by newspapers or periodical agencies which require information in different fields such as economic trends, business, stock and share markets, sports, politics and so on.

NOTES

Merits

- This method is very cheap and economical for extensive investigations.
- The required information can be obtained expeditiously since only rough estimates are required.

Limitations

Since the correspondents apply their own judgement about the method of collecting the information, the results are often vitiated due to personal prejudices and whims of the correspondents. The data so obtained is thus not so reliable. This method is suitable only if the purpose of investigation is to obtain rough and approximate estimates. It is unsuited where a high degree of accuracy is desired.

(c) Experiments

Experiments are also a method of collecting data. Experiments are resorted to when it is required to collect factual data when nothing is available for reference. It may also be conducted to verify the theory. It is a study conducted under controlled conditions. Experiments are made by researchers to understand the cause and effect relationships. Such relationships are also made in observational studies but here, there is no control on how subjects are assigned to groups.

Experimental design

This design contains information gathering exercises that have variations under control of the experimenter. In observational studies, there is no control on condition. Mostly, an experimenter wants to know the effect of some process on certain objects, which are taken as 'experimental units'. Such objects are either a small section of people, few groups, etc. Such design finds broad application in natural and social sciences.

An experiment was carried by a surgeon to develop a cure for scurvy. The experiment was conducted in controlled conditions. He selected a dozen of men who were travelling in a ship. All 12 men had scurvy. He made six pairs from these 12 people and prescribed them diet for two weeks and each was given a different one. These were as follows:

- One fourth of cider daily
- Sulphuric acid (twenty five drops), thrice a day on empty stomach
- Seawater, half-pint daily
- Garlic, mustard, horseradish and mixed together
- Vinegar, two teaspoon, full, three times daily
- Lemon, one and oranges two, daily.

A condition of every pair was noted and it was found that the pair that consumed citrus fruits recovered fast and reported for duty after 6 days. Others too recovered but the rate of recovery was not comparable to this group.

In this study, he restricted sample size to avoid variation for extraneous reasons. The paired men provided replication. The main thing missing is randomized allocation of subjects for medical treatment.

Analysis of the experimental design has the foundation of variance analysis. This analysis is done by collecting models having variance already observed, and these were partitioned into different components on different factors, and then estimation and testing were carried out.

We now consider another experiment where eight number if objects is to be weighed using a pan balance and a set of few standard weights. Each instrument weighs the difference between objects in the left pan verses those in the right pan. Further, there is an addition of standard weights that were kept on the lighter pan and equilibrium point is noted. There was a random error for each experiment averaging zero. Standard deviation errors, due to the probability distribution, are σ on different weights and these are independent. We denote true weight as θ_1, ..., θ_8.

Experiments considered are,

1. Weighing of each object in one pan, while the other is empty. We denote X_i as the weight of the ith object, where i vary from 1 to 8.
2. Carry on weighing of eight as per schedule given below. We take measured difference as Y_i where i vary from 1 to 8.

	Left pan	Right pan
1st weighing:	1 2 3 4 5 6 7 8	(empty)
2nd:	1 2 3 8	4 5 6 7
3rd:	1 4 5 8	2 3 6 7
4th:	1 6 7 8	2 3 4 5
5th:	2 4 6 8	1 3 5 7
6th:	2 5 7 8	1 3 4 6
7th:	3 4 7 8	1 2 5 6
8th:	3 5 6 8	1 2 4 7

The weight θ_1 has estimated value of

$$\hat{\theta}_1 = \frac{Y_1 + Y_2 + Y_3 + Y_4 - Y_5 - Y_6 - Y_7 - Y_8}{8}.$$

Estimated value for weights of the other items, θ_2 is

$$\hat{\theta}_2 = \frac{Y_1 + Y_2 - Y_3 - Y_4 + Y_5 + Y_6 - Y_7 - Y_8}{8}.$$

In decision-making one has to choose better alternative. Here, σ^2 is the variance of the estimate X_1 of θ_1 for first experiment. But $\sigma^2/8$ is the variance for second experiment. Thus, there is 8 times more precision in the second experiment, for a single item. Estimates are done for all simultaneously having same precision. If weighed separately 64 weight are to taken with 8 weighing in second experiment. Estimates for items in second experiment have errors, correlated to other.

NOTES

This also serves as an example for the design of experiments that involve combinatorial designs.

(i) **Select problem:** For designing an experiment, one must select a problem and put phrase for it. This will direct the design as well as outcomes of the experiment. Issues related to questions like 'Who, What, When, Why and How' need to be addressed. Let us take the case on automobile accidents and design an experiment for this. We now collect data for this experiment. Depending on presentation of the problem is stated, aim of the experiment may be different. This may either lead to the design of a road surface for existing automobiles or a brand new automobile. To make research more precise, covering greater depth, proven models should underlie in making a design for the experiment.

(ii) **Determining dependent variables:** Dependent variables need measurement in the experiment and these dependent variables may be various. Variables should be split into system level and individual level. Questions are on the experiments only when taken for a system level. Such variables are created so that a conclusion can be drawn. Further, such conclusions should be supported from as many different angles as possible. Such operations are called converging operations. System level dependent variable tells as to how many experimenters are there while a certain task is being done. If taken at individual level, these dependent variables are taken as measurements for a particular subject. Such measurements of dependent variables are to be analyzed and reduced.

Dependent variables may consist of different measures like performance, subjective and physical. Performance measures tell time taken by the participant in completing the task plus number of mistakes made during the task. Subjective measures tell about the method used or not used by participants.

(iii) **Determining independent variables:** These variables get manipulated in experiment. These are the things related to people and these are typically, sex, age, level of education, general work experience or vision. To ensure meeting of specifications, subjects are to be screened prior to running the experiment.

(iv) **Determining the number of levels of independent variables:** This determines as to how much number of experimental conditions is to be manipulated. If an experiment is to be designed for assessing relative performance of few automobiles, say 10, then independent variables have number of levels as 10.

(v) **Determining the possible combinations:** There is a need to establish types of combinations in independent variables. Only then an experiment can be taken as valid. In earlier example on automobiles, comparison of Model need not be considered feasible. Let us assume that model 'A' has automatic transmission whereas Model 'B' has manual transmission. Establishing all possible combination is important here.

(vi) **Determining the number of observations:** Depending on desired analysis, certain factors are to be considered before deciding on number of observations. This consists of number as to how many trials are to be taken to familiarize with the experiment.

(vii) **Redesign:** This is necessary for obtaining the optimal design. Redesign is essential when there are certain lacunas in the experiment design. Inconsistencies are given by inaccuracy while stating the problem, selection of inadequate variables and non availability of desired apparatus. Recommended timeframe for redesign is:

- Planning and scheduling 44 per cent
- Testing 6–10 per cent
- Reduction, analysis and writing 45–50 per cent

(viii) **Randomization:** It is a trial that is randomized and controlled is most reliable and impartial method. It is a process of assigning participants not by choice, but by chance. This is done either to the group carrying out investigation or those who are controlling. This ensures trials do not receive the preferred results.

(ix) **Data collection:** The data collection must ensure that these experiments are supported by factual data. This lies in collection of raw data and but adhering to the experimental conditions. The data here may be very large.

(x) **Data reduction:** For data reduction, raw data are taken into manageable chunks, for further utilization. Entire data may not be found pertinent and thus need to be excluded and need not be considered for analysis.

(xi) **Data verification:** This is essential and mostly carried out by plotting reduced data that gives a visual picture of how data is located. These points indicate erroneous data collection.

True experimental design needs an environment that is created for control of spurious data that may mislead the experimental conclusion. A purchase laboratory makes an approach, most suited for this. Researchers modify one variable at a time to determine the effect on sales volume. Virtual purchase labs, which are Internet-based labs, are becoming popular.

Classification of Data

When the raw data has been collected and edited, it should be put into an ordered form (ascending or descending order), so that it can be looked at more objectively. The next important step towards processing the data is classification. Classification means separating items according to similar characteristics and grouping them into various classes. The items in different classes will differ from each other on the basis of some characteristics or attributes. Classification of data is very similar to sorting of mail at a post office, where a mail is classified according to its geographical destination and may further be classified into the type of mail such as first class, parcel post, and so on. The data may be classified into four broad classes:

(*i*) **Geographical.** This classification groups the data according to locational differences among the items. The geographical areas are usually listed in

NOTES

alphabetical order for easy reference. For example, the book listing colleges and universities in various states in India would first list the states in the alphabetical order and then the colleges and the universities within these states in the alphabetical order.

(*ii*) **Chronological.** This classification includes data according to the time period in which the items under consideration occurred. For example, the sales of automobiles in India over the last ten years may be grouped by to the year in which such sales took place.

(*iii*) **Qualitative.** In this type of classification, the data is grouped together according to some distinguished characteristic or attribute such as religion, sex, age, national origin, and so on. This classification simply identifies whether a given attribute is present or absent in a given population. For example, the population may be divided into two classes: males and females. Then the attribute of male will go into one class and the attribute of female will go into the other.

(*iv*) **Quantitative.** This refers to the classification of data according to some attribute which has magnitude and can be measured such as classification according to weight, height, income, and so on. For example, the salaries of professors at a university may be classified according to their rank such as instructor, assistant professor, associate professor and full professor.

Hence, the collected data should be arranged systematically to give it shape, form and meaning. The division of the data into homogeneous groups according to their characteristics, recorded in a statistical inquiry, is called *classification.*

Primary and Secondary Data

The statistical data, as previously discussed, may be classified under two categories depending upon the sources utilized. These categories are:

(i) **Primary data.** Primary data is one which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by surveys conducted by individuals or research institutions. For example, if a researcher is interested to know what women think about the issue of abortion, he/she must undertake a survey and collect data on the opinions of women by asking relevant questions. Such data collected would be considered as primary data.

(ii) **Secondary data.** When an investigator uses the data which has already been collected by others, such data is called secondary data. This data is primary data for the agency that collected it and becomes secondary data for someone else who uses this data for his own purposes. The secondary data can be obtained from journals, reports, government publications, publications of professional and research organizations and so on. For example, if a researcher desires to analyse the weather conditions of different regions, he can get the required information or data from the records of the meteorology department. Even though secondary data is less expensive to collect in terms of money and time, the quality of this data may even be better under certain situations, because it may have been collected by persons who were specifically trained for that purpose. However, such secondary data must be used with utmost care. The reason is that such data may be full of errors due to the fact that the purpose of the collection of data by the primary agency may have been different from that of the user of the secondary data. Additionally, there may have been biases introduced during collection of data or analysis of data. For example, the size of the sample may have been inadequate or there may have been arithmetical or definitional errors. Hence, it is necessary to critically investigate the validity of the secondary data as well as the credibility of the primary data collection agency.

Check Your Progress

3. What is direct personal observation?
4. How are questions prepared for a mailed questionnaire?
5. What are the merits of interview method?
6. What are the limitations of telephone survey method?
7. Why are experiments conducted?

1.4 PRESENTATION OF DATA

In this section, we will study about different methods of presenting data.

Given below are the methods of presenting data along with their types:

Tables	Graphs
• General tables or Reference tables	• Line Graph
• Specific purpose or Derivative tables	• Frequency Polygon
• Simple and Complex tables	• Ogive curves
	• Histograms
	• Diagrams
	• Pie Diagram

(i) Tables

Classification of data is usually followed by tabulation, which is considered as the mechanical part of classification.

Tabulation is the systematic arrangement of data in columns and rows. The analysis of the data is done so by arranging the columns and rows to facilitate analysis and comparisons.

Tabulation has the following objectives:

(*i*) Simplicity. The removal of unnecessary details gives a clear and concise picture of the data

(*ii*) Economy of space and time

(*iii*) Ease in comprehension and remembering

(*iv*) Facility of comparisons. Comparisons within a table and with other tables may be made

(*v*) Ease in handling of totals, analysis, interpretation, etc.

Construction of Tables

A table is constructed depending on the type of information to be presented and the requirements of statistical analysis. The following are the essential features of a table:

NOTES

(*i*) *Title*. It should have a clear and relevant *title*, which describes the contents of the table. The title should be brief and self explanatory.

(*ii*) *Stubs and captions*. It should have clear headings and sub headings. Column headings are called *captions* and row headings are called *stubs*. The stubs are usually wider than the captions.

(*iii*) *Unit*. It should indicate all the *units* used.

(*iv*) *Body*. The *body* of the table should contain all information arranged according to description.

(*v*) *Headnote*. The *headnote* or prefatory note, placed just below the title, in a less prominent type, gives some additional explanation about the table. Sometimes, the headnote consists of the unit of measurement.

(*vi*) *Footnotes*. A *footnote* at the bottom of the table may clarify some omissions of special features.

A source note gives information about the source used, if any.

(*vii*) *Arrangement of data*. Data may be arranged according to requirements in chronological, alphabetical, geographical, or any other order.

(*viii*) *Emphasis*. The items to be emphasized may be put in different print or marked suitably.

(*ix*) *Other details*. Percentages, ratios, etc. should be shown in separate columns. Thick and thin lines should be drawn at proper places.

A table should be easy to read and should contain only the relevant details. If the aim of clarification is not achieved, the table should be redesigned.

Types of Tables

Depending on the nature of the data and other requirements, tables may be divided into various types.

- *General tables or Reference tables*. These contain detailed information for general use and reference, e.g., tables published by government agencies.
- *Specific purpose or Derivative tables*. They are usually summarized from general tables and are useful for comparison and analytical purposes. Averages, percentages etc. are incorporated along with information in these tables.
- *Simple and Complex tables*. A table showing only one characteristic is a simple table (see Table 1.1). The more common tables are complex and show two or more characteristics or groups of items.

Table 1.1 *Simple Table*

Cinema Attendance among Adult Male Factory Workers in Bombay
March 1972

Frequency	*Number of Workers*
Less than once a month	3780
1 to 4 times a month	1652
More than 4 times a month	926

Table 1.2 is the result of a survey on the cinema going habits of adult factory workers.

Table 1.2 *Simple Table*

Cinema Attendance among Adult Male Factory Workers in Bombay

March 1972

Cinema Attendance	*Single*		*Married*	
Frequency	*Under 30*	*Over 30*	*Under 30*	*Over 30*
Less than once a month	122	374	1404	1880
1–4 times a month	1046	202	289	115
More than 4 times a month	881	23	112	10
Total	2049	599	1805	2005

It is obvious that the tabular form of classification of data is a great improvement over the narrative form.

Frequently, table construction involves deciding which attribute should be taken as primary and which as secondary. For the previous table, we can also consider that whether it would be improved further if 'under 30' and '30 and over' had been the main column headings and 'single' and 'married' the sub headings. The modifications depend on the purpose of the table. If the activities of *age groups* are to be compared, it is best left as it stands. But if a comparison between men of different *marital status* is required, the change would be an improvement.

Advantages of Tabulation of Data

(*i*) Tabulated data can be more easily understood and grasped than untabulated data.

(*ii*) A table facilitates comparisons between subdivisions and with other tables.

(*iii*) It enables the required figures to be located easily.

(*iv*) It reveals patterns within the figures, which otherwise might not have been obvious, e.g., from the previous table, we can conclude that regular and frequent cinema attendance is mainly confined to younger age group.

(*v*) It makes the summation of items and the detection of errors and omissions easier.

(*vi*) It obviates repetition of explanatory phrases and headings and hence takes less space.

(ii) Graphs

In a graph, the independent variable should always be placed on the horizontal or x-axis and the dependent variable on the vertical or y-axis.

(a) Line Graph

Here, the points are plotted on paper (or graph paper) and joined by straight lines. Generally, continuous variables are plotted by line graph.

Example 1.1: The monthly averages of Retail Price Index from 1996 to 2003 (Jan. 1996 = 100) were as follows:

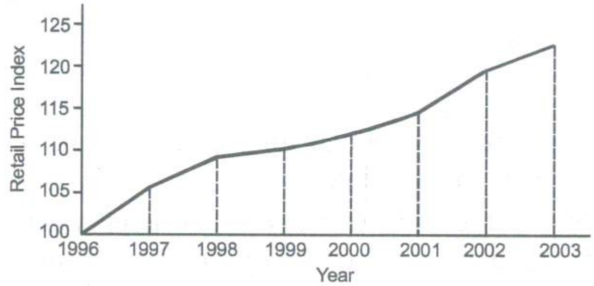
Year	1996	1997	1998	1999	2000	2001	2002	2003
Retail Price Index	100	105.8	109.0	109.6	110.7	114.5	119.3	122.3

NOTES

Draw a diagram to display these figures.

Solution: Here, years are plotted along the horizontal line and the retail price index along the vertical line.

Erect perpendiculars to horizontal line from the points marked as retail price index for the years 1997, 1998, ..., 2003 and cut off these ordinates according to the given data and thus various points will be plotted on the paper. Join these points by straight lines.



(b) Frequency Polygon

A frequency polygon is a line chart of frequency distribution in which, either the values of discrete variables or midpoints of class intervals are plotted against the frequencies and these plotted points are joined together by straight lines. Since the frequencies generally do not start at zero or end at zero, this diagram as such would not touch the horizontal axis. However, since the area under the entire curve is the same as that of a histogram which is 100 per cent of the data presented, the curve can be enclosed so that the starting point is joined with a fictitious preceding point whose value is zero. This ensures that the start of the curve is at horizontal axis and the last point is joined with a fictitious succeeding point whose value is also zero, so that the curve ends at the horizontal axis. This enclosed diagram is known as the frequency polygon.

We can construct the frequency polygon from the table presented for the ages of 30 workers as follows:

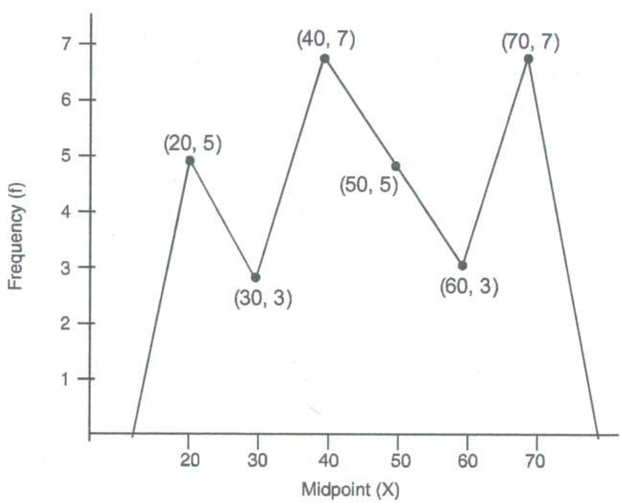


Fig. 1.1 Frequency Polygon

NOTES

(c) Relative Frequency

In a frequency distribution, if the frequency in each class interval is converted into a proportion, dividing it by the total frequency, we get a series of proportions called *relative frequencies*. A distribution presented with relative frequencies rather than actual frequencies is called a *relative frequency* distribution. The sum of all relative frequencies in a distribution is 1.

Example 1.2: Calculate relative frequency from the given table.

Class Interval	*Frequency*
25—35	7
35—45	9
45—55	22
55—65	7
65—75	3
75—85	2

Solution: This example shows that the sum of all relative frequencies in a distribution is 1.

Class Interval	*Frequency*	*Relative Frequency*	*Explanation*
25—35	7	0.14	$\frac{7}{50} = 0.14$
35—45	9	0.18	$\frac{9}{50} = 0.18$
45—55	22	0.44	etc.
55—65	7	0.14	
65—75	3	0.06	
75—85	2	0.04	
	Total 50	1.00	

The concept of relative frequencies is useful in sampling theory. It can also be used to compare two frequency distributions with unequal total frequency with the same series of class intervals as in the following example.

Example 1.3: Compare the following frequency distribution.

Class Interval	f_1	f_2
10—20	5	12
20—30	10	24
30—40	6	30
40—50	3	19
50—60	1	15

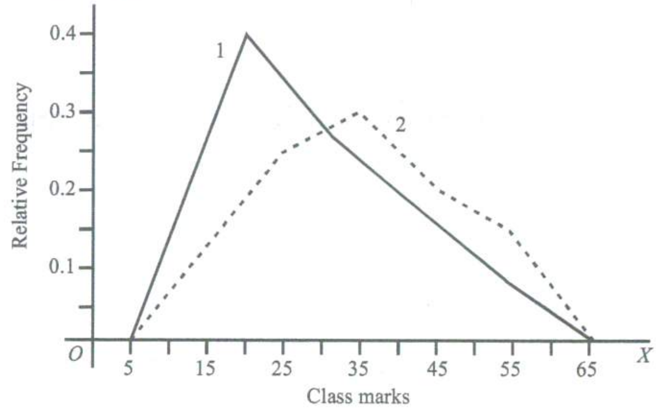
Solution: The following table shows the comparison.

Class Interval	f_1	f_2	*Rel. Freq.* f_1	*Rel. Freq.* f_2
10—20	5	12	0.20	0.12
20—30	10	24	0.40	0.24
30—40	6	30	0.24	0.30
40—50	3	19	0.12	0.19
50—60	1	15	0.04	0.15
Total	25	100	1.00	1.00

A direct visual comparison of two frequency distributions can be made by drawing their frequency polygons.

Example 1.4: Draw frequency polygons for the relative frequency distributions given in example 1.3.

Solution: The following is the frequency polygon for the relative frequencies as mentioned in example 1.3.



(d) Ogive Curves

Cumulative frequency curve or ogive is the graphic representation of a cumulative frequency distribution. Ogives are of two types. One of these is less than and the other one is greater than ogive. Both these ogives are constructed based upon the following table of our example of 30 workers.

Table 1.3 *Cumulative Frequency Distribution*

Class Interval (years)	*Mid-point*	*(f)*	*Cum. Freq. (less than)*	*Cum. Freq. (greater than)*
15 and upto 25	20	5	5 (less than 25)	30 (more than 15)
25 and upto 35	30	3	8 (less than 35)	25 (more than 25)
35 and upto 45	40	7	15 (less than 45)	22 (more than 35)
45 and upto 55	50	5	20 (less than 55)	15 (more than 45)
55 and upto 65	60	3	23 (less than 65)	10 (more than 55)
65 and upto 75	70	7	30 (less than 75)	7 (more than 65)

(*i*) **Less than ogive.** In this case, the less than cumulative frequencies are plotted against the upper boundaries of their respective class intervals.

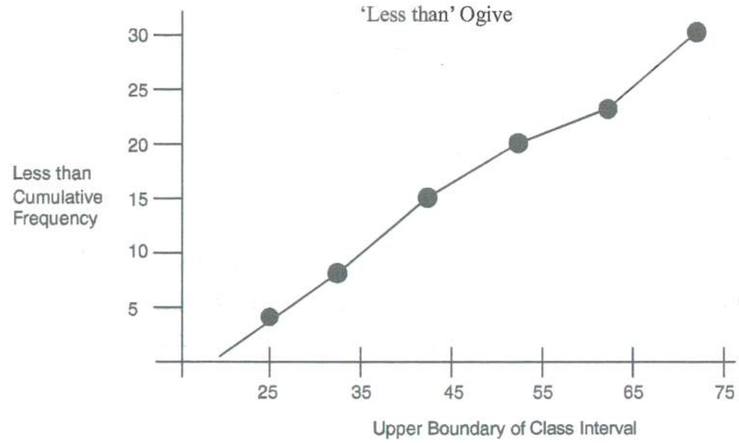


Fig. 1.2 'Less than' Ogive

(*ii*) **Greater than ogive.** In this case, the greater than cumulative frequencies are plotted against the lower boundaries of their respective class intervals.

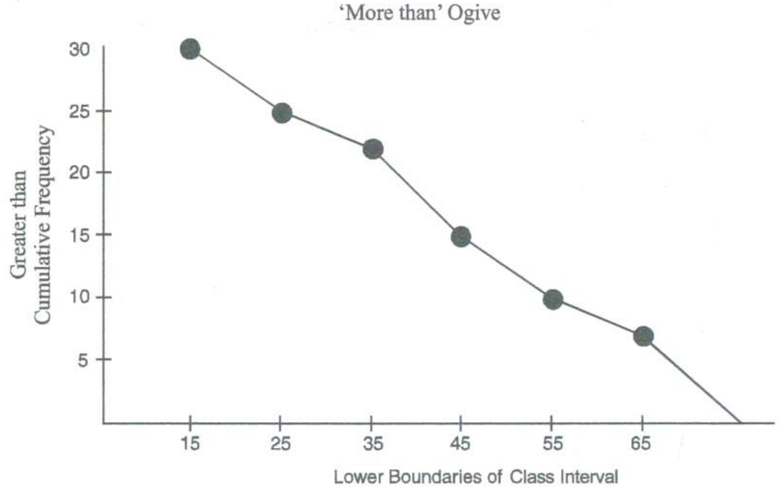


Fig. 1.3 'More than' Ogive

These ogives can be used for comparison purposes. Several ogives can be drawn on the same grid, preferably with different colours for easier visualization and differentiation.

Although, diagrams and graphs are powerful and effective media for presenting statistical data, they can only represent a limited amount of information and they are not of much help when intensive analysis of data is required.

(e) Histograms

A histogram is the graphical description of data and is constructed from a frequency table. It displays the distribution method of a data set and is used for statistical as well as mathematical calculations.

NOTES

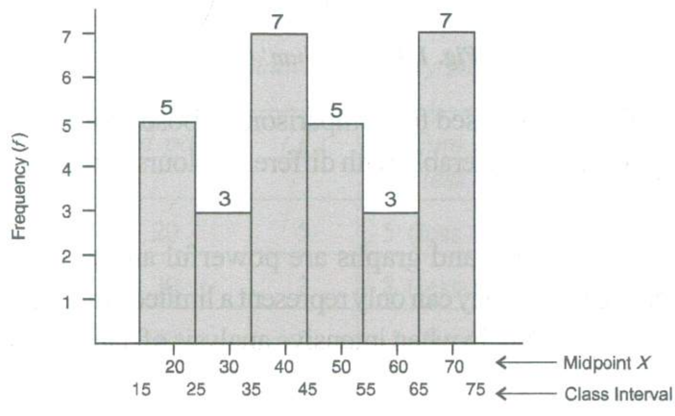
The word histogram is derived from the Greek word *histos* which means 'anything set upright' and *gramma* which means 'drawing, record, writing'. It is considered as the most important basic tool of statistical quality control process.

In this type of representation, the given data is plotted in the form of a series of rectangles. Class intervals are marked along the X-axis and the frequencies along the Y-axis according to a suitable scale. Unlike the bar chart, which is one dimensional, meaning that only the length of the bar is important and not the width, a histogram is two-dimensional in which both the length and the width are important. A histogram is constructed from a frequency distribution of a grouped data, where the height of the rectangle is proportional to the respective frequency and the width represents the class interval. Each rectangle is joined with the other and any blank spaces between the rectangles would mean that the category is empty and there are no values in that class interval.

As an example, let us construct a histogram for our example of ages of 30 workers. For convenience sake, we will present the frequency distribution along with the midpoint of each interval, where the midpoint is simply the average of the values of the lower and the upper boundary of each class interval. The frequency distribution table is shown as follows:

Class Interval (years)	*Midpoint*	*(f)*
15 and upto 25	20	5
25 and upto 35	30	3
35 and upto 45	40	7
45 and upto 55	50	5
55 and upto 65	60	3
65 and upto 75	70	7

The histogram of this data would be shown as follows:



(iii) Diagrams

The data we collect can often be more easily understood for interpretation if it is presented graphically or pictorially. Diagrams and graphs give visual indications of magnitudes, groupings, trends and patterns in the data. These important features

NOTES

are more simply presented in the form of graphs. Also, diagrams facilitate comparisons between two or more sets of data.

The diagrams should be clear and easy to read and understand. Too much information should not be shown in the same diagram; otherwise, it may become cumbersome and confusing. Each diagram should include a brief and self explanatory title dealing with the subject matter. The scale of the presentation should be chosen in such a way that the resulting diagram is of appropriate size. The intervals on the vertical as well as the horizontal axis should be of equal size; otherwise, distortions would occur.

Diagrams are more suitable to illustrate the data which is discrete, while continuous data is better represented by graphs. The following are the diagrammatic and the graphic representation methods that are commonly used.

(a) One Dimensional Diagrams

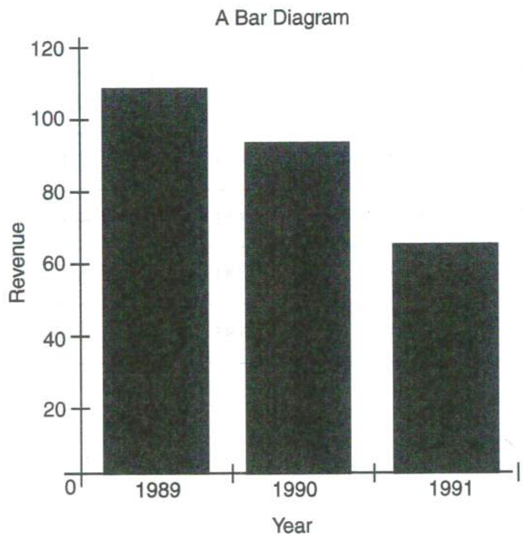
Bars are simply vertical lines where the lengths of the bars are proportional to their corresponding numerical values. The width of the bar is unimportant but all bars should have the same width so as not to confuse the reader of the diagram. Additionally, the bars should be equally spaced.

Example 1.5: Suppose that the following were the gross revenues (in $100,000.00) for a company *XYZ* for the years 1989, 1990 and 1991.

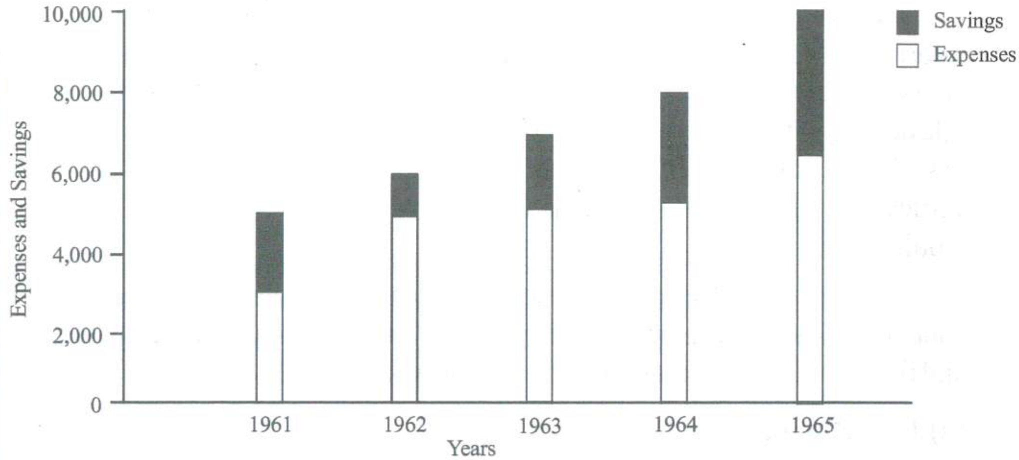
Year	*Revenue*
1989	110
1990	95
1991	65

Construct a bar diagram for this data.

Solution: The bar diagram for this data can be constructed as follows with the revenues represented on the vertical axis and the years represented on the horizontal axis.



When each figure is made up of two or more component figures, the bars may be subdivided into components. Too many components should not be shown.



Component Bar Chart Showing Expenses and Savings of Mr X

Annual Income, Expenses and Savings of Mr X

Year	*Amounts in ₹ of*			*Percentages of*		
	Income	*Expenses*	*Savings*	*Income*	*Expenses*	*Savings*
1961	5000	3000	2000	100.0	60.0	40.0
1962	6000	5000	1000	100.0	83.3	16.7
1963	7000	5000	2000	100.0	71.4	28.6
1964	8000	5000	3000	100.0	62.5	37.5
1965	10000	6000	4000	100.0	60.0	40.0

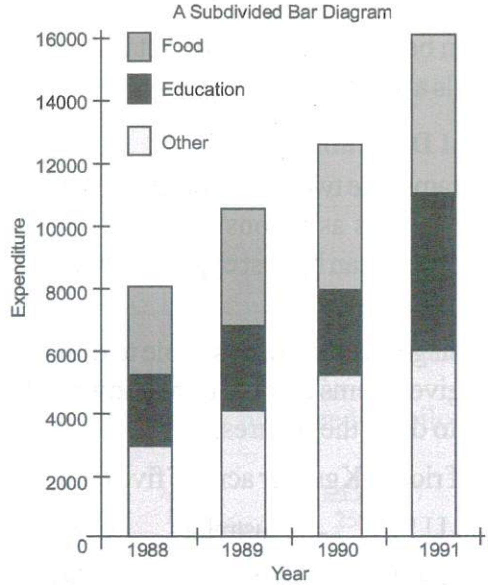
The bars drawn can be further subdivided into components depending upon the type of information to be shown in the diagram. This will be clear by the following example in which we present three components in a bar.

Example 1.6: Construct a subdivided bar chart for the three types of expenditures in dollars for a family of four for the years 1988, 1989, 1990 and 1991 given as follows:

Year	*Food*	*Education*	*Other*	*Total*
1988	3000	2000	3000	8000
1989	3500	3000	4000	10500
1990	4000	3500	5000	12500
1991	5000	5000	6000	16000

NOTES

Solution: The subdivided bar chart would be as follows:



Percentage component bars or divided bar charts

When in the previous case, the component lengths represent the percentages (instead of the actual amounts) of each component we get percentage component bar charts. The heights of all the bars will be the same.

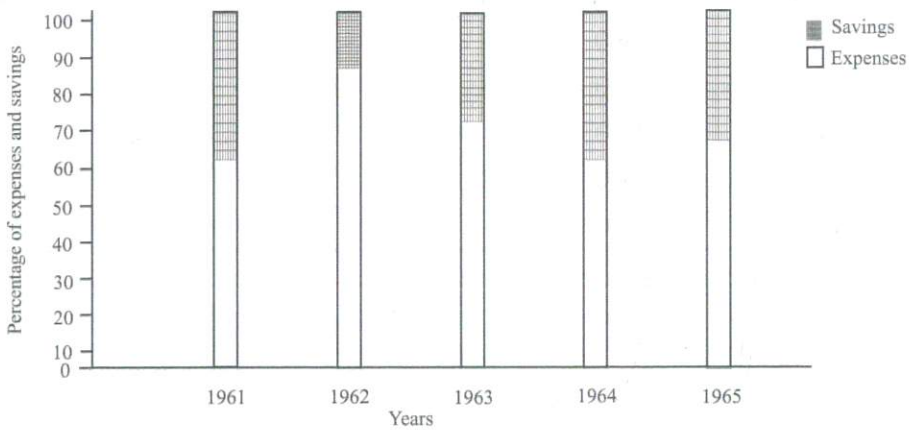


Fig. 1.4 Percentage Component Bar Chart showing Expenses and Savings of Mr X

Multiple Bar Charts

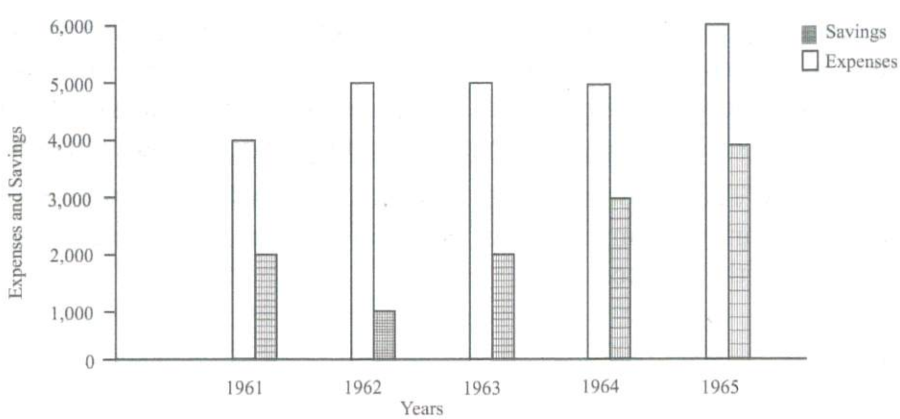


Fig. 1.5 Multiple Bar Chart showing Expenses and Savings of Mr X

NOTES

Here, the interrelated component parts are shown in adjoining bars, coloured or marked differently, thus allowing comparison between different parts.

These charts can be used if the overall total is not required. Some charts given earlier show totals also.

(b) Two Dimensional Diagrams

Two dimensional diagrams take two components of data for representation. These are also called area diagrams as it considers two dimensions. The types are rectangles, squares and pie. It can be best explained with the help of the following squares diagram example:

Squares: The square diagram is easy and simple to draw. Take the square root of the values of various given items that are to be shown in the diagrams and then select a suitable scale to draw the squares.

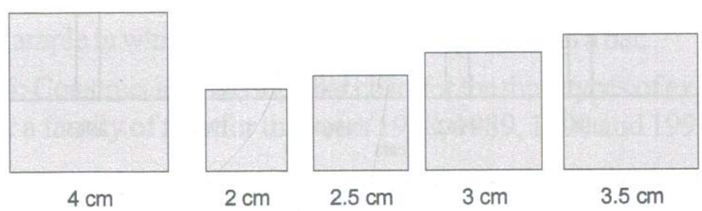
Example 1.7: Yield of rice in Kgs. per acre of five countries are as follows:

Country	U.S.A.	Australia	U.K.	Canada	India
Yield of rice in Kgs. per acre	6400	1600	2500	3600	4900

Represent this data using square diagram.

Solution: To draw the square diagrams calculate as follows:

Country	*Yield*	*Square root*	*Side of the square in cm*
U.S.A	6400	80	4
Australia	1600	40	2
U.K.	2500	50	2.5
Canada	3600	60	3
India	4900	70	3.5



(c) Pie Diagram

This type of diagram enables us to show the partitioning of a total into its component parts. The diagram is in the form of a circle and is also called a pie because the entire diagram looks like a pie and the components resemble slices cut from it. The size of the slice represents the proportion of the component out of the whole.

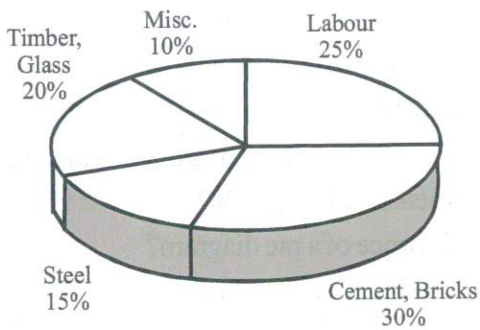
Example 1.8: The following figures relate to the cost of the construction of a house. The various components of cost that go into it are represented as percentages of the total cost.

Item	*% Expenditure*
Labour	25
Cement, bricks	30
Steel	15
Timber, glass	20
Miscellaneous	10

NOTES

Construct a pie chart for the above data.

Solution: The pie chart for this data is presented as follows:



Pie charts are very useful for comparison purposes, especially when there are only a few components. If there are too many components, it may become confusing to differentiate the relative values in the pie.

(d) Three Dimensional Diagrams

Three dimensional diagrams are also termed as volume diagram and consist of cubes, cylinders, spheres, etc. In these diagrams, three dimensions namely length, width and height are taken into account. Cubes are used to represent where side of a cube is drawn in proportion to the cube root of the magnitude of data.

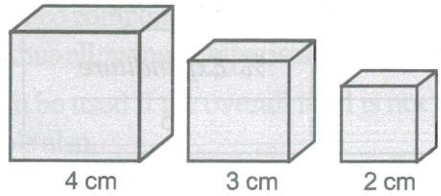
Example 1.9: Represent the following data using volume diagram.

Category	*Number of Students*
Under graduate	64000
Post graduate	27000
Professionals	8000

Solution: The sides of cubes are calculated as follows:

Category	*Number of Students*	*Cube Root*	*Side of Cube*
Undergraduate	64000	40	4 cm
Postgraduate	27000	30	3 cm
Professional	8000	20	2 cm

NOTES



Check Your Progress

8. Define tabulation.
9. What are the objectives of tabulation?
10. What do you understand by line graph?
11. What is frequency polygon?
12. What do you mean by the term 'relative frequency'?
13. What are ogive curves?
14. State the types of ogives.
15. What are histograms?
16. What kinds of data are illustrated by graphs and diagrams?
17. What are two dimensional diagrams?
18. What is the significance of a pie diagram?
19. What are three dimensional diagrams?

1.5 BASIC TOOLS OF DATA ANALYSIS

Data analysis tools are defined as a series of maps, diagrams and charts designed to collate, analyse and present data for a wide range of applications and industries. Various programs and methodologies have been developed for use in all types of industry, ranging from manufacturing and quality assurance to research groups and data collection companies. The tools of data analysis are used to classify software and applications used by data analysts to create and execute analytic processes that help businesses make smarter, more informed business decisions while minimising cost and maximising profits.

1.5.1 Measures of Central Tendency

This section pertains to mean, median, mode, quartiles, deciles, percentiles, measures of dispersion, standard deviation and coefficient of variation.

(i) Arithmetic Mean

There are several commonly used measures such as arithmetic mean, mode and median. These values are very useful not only in presenting the overall picture of the entire data but also for the purpose of making comparisons among two or more sets of data.

As an example, questions like 'How hot is the month of June in Delhi?' can be answered generally by a single figure of the average for that month. Similarly, suppose we want to find out if boys and girls of age 10 years differ in height for the purpose of making comparisons. Then, by taking the average height of boys of that age and the average height of girls of the same age, we can compare and record the differences.

NOTES

While arithmetic mean is the most commonly used measure of central tendency, mode and median are more suitable measures under certain set of conditions and for certain types of data. However, each measure of central tendency should meet the following requisites.

- It should be easy to calculate and understand.
- It should be rigidly defined. It should have only one interpretation so that the personal prejudice or the bias of the investigator does not affect its usefulness.
- It should be representative of the data. If it is calculated from a sample, the sample should be random enough to be accurately representing the population.
- It should have a sampling stability. It should not be affected by sampling fluctuations. This means that if we pick ten different groups of college students at random and compute the average of each group, then we should expect to get approximately the same value from each of these groups.
- It should not be affected much by extreme values. If few, very small or very large items are present in the data, they will unduly influence the value of the average by shifting it to one side or other, so that the average would not be really typical of the entire series. Hence, the average chosen should be such that it is not unduly affected by such extreme values.

Arithmetic mean is also commonly known as the mean. Even though average, in general, means measure of central tendency, when we use the word average in our daily routine, we always mean the arithmetic average. The term is widely used by almost everyone in daily communication. We speak of an individual being an average student or of average intelligence. We always talk about average family size or average family income or grade point average (GPA) for students, and so on.

For discussion purposes, let us assume a variable X which stands for some scores such as the ages of students. Let the ages of 5 students be 19, 20, 22, 22 and 17 years. Then variable X would represent these ages as follows:

$$X: 19, 20, 22, 22, 17$$

Placing the Greek symbol Σ(Sigma) before X would indicate a command that all values of X are to be added together. Thus:

$$\Sigma X = 19 + 20 + 22 + 22 + 17$$

The mean is computed by adding all the data values and dividing it by the number of such values. The symbol used for sample average is $\overline{X}$ so that:

$$\overline{X} = \frac{19 + 20 + 22 + 22 + 17}{5}$$

In general, if there are n values in the sample, then

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

In other words,

NOTES

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad i = 1,\ 2\ ...\ n$$

According to this formula, the mean can be obtained by adding up all values of X_i, where the value of i starts at 1 and ends at n with unit increments so that i = 1, 2, 3, ... n.

If instead of taking a sample, we take the entire population in our calculations of the mean, then the symbol for the mean of the population is μ(mu) and the size of the population is N, so that:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}, \qquad i = 1,\ 2\ ...N$$

If we have the data in grouped discrete form with frequencies, then the sample mean is given by:

$$\overline{X} = \frac{\Sigma f(X)}{\Sigma f}$$

Here, Σf = Summation of all frequencies
= n

$\Sigma f(X)$ = Summation of each value of X multiplied by its corresponding frequency (f)

Example 1.10: Let us take the ages of 10 students as follows:

19, 20, 22, 22, 17, 22, 20, 23, 17, 18

Solution: This data can be arranged in a frequency distribution as follows:

(X)	(f)	f(X)
17	2	34
18	1	18
19	1	19
20	2	40
22	3	66
23	1	23
	Total = 10	200

In this case, we have Σf = 10 and $\Sigma f(X)$ = 200, so that:

$$\overline{X} = \frac{\Sigma f(X)}{\Sigma f}$$

$$= 200/10 = 20$$

Example 1.11: Calculate the mean of the marks of 46 students given in the following table.

NOTES

Frequency of Marks of 46 Students

Marks (X)	*Frequency* (f)
9	1
10	2
11	3
12	6
13	10
14	11
15	7
16	3
17	2
18	1
Total	46

Solution: This is a discrete frequency distribution, and is calculated using the equation $\overline{x} = \frac{\Sigma f(x)}{\Sigma f}$. The following table shows the method of obtianing $\Sigma f(X)$.

Marks (X)	*Frequency* (f)	$f(X)$
9	1	9
10	2	20
11	3	33
12	6	72
13	10	130
14	11	154
15	7	105
16	3	48
17	2	34
18	1	18
	$\Sigma f = 46$	$\Sigma f(X) = 623$

$$\overline{X} = \frac{\Sigma f(X)}{\Sigma f} = \frac{623}{46} = 13.54$$

Characteristics of the mean

The arithmetic mean has three interesting properties. These are as follows:

(i) The sum of the deviations of individual values of X from the mean will always add up to zero. This means that if we subtract all the individual values from their mean, then some values will be negative and some will be positive, but if all these differences are added together then the total sum will be zero. In other words, the positive deviations must balance the negative deviations. Or symbolically:

$$\sum_{i=1}^{n}(X_i - \overline{X}) = \quad 0, i = 1, 2, \ldots n$$

(ii) The second important characteristic of the mean is that it is very sensitive to extreme values. Since the computation of the mean is based upon inclusion of all values in the data, an extreme value in the data would shift the mean towards it, thus making the mean unrepresentative of the data.

(iii) The third property of the mean is that the sum of squares of the deviations about the mean is minimum. This means that if we take differences between individual values and the mean and square these differences individually and then add these squared differences, then the final figure will be less than the sum of the squared deviations around any other number other than the mean. Symbolically, it means that:

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \text{Minimum}, i = 1, 2, \ldots n$$

The following examples will make the concept clear about properties of mean.

(iv) The product of the arithmetic mean and the number of values on which the mean is based is equal to the sum of all given values. In other words, if we replace each item in series by the mean, then the sum of the these substitutions will equal the sum of individual items. Thus, in the figures 3, 5, 7, 9, if we substitute the mean for each item 6, 6, 6, 6 then the total is 24, both in the original series and in the substitution series.

This can be shown like this.

Since, $\overline{X} = \frac{\Sigma X}{N}$

$\therefore$ $N\overline{X} = \Sigma X$

For example, if we have a series of values 3, 5, 7, 9, the mean is 6. The squared deviations are:

X	$X - \overline{X} = X'$	X'^2
3	$3-6=-3$	9
5	$5-6=-1$	1
7	$7-6=1$	1
9	$9-6=3$	9
		$\Sigma X'^2 = 20$

This property provides a test to check if the computed value is the correct arithmetic mean.

Example 1.12: The mean age of a group of 100 persons (grouped in intervals 10–, 12–,..., etc.) was found to be 32.02. Later, it was discovered that age 57 was misread as 27. Find the corrected mean.

Solution: Let the mean be denoted by *X*. So, putting the given values in the formula of arithmetic mean, we have,

$$32.02 = \frac{\sum X}{100}, \text{ i.e., } \sum X = 3202$$

$$\text{Correct } \sum X = 3202 - 27 + 57 = 3232$$

$$\therefore \quad \text{Correct AM} = \frac{3232}{100} = 32.32$$

Example 1.13: The mean monthly salary paid to all employees in a company is ₹ 500. The monthly salaries paid to male and female employees average ₹ 520 and ₹ 420, respectively. Determine the percentage of males and females employed by the company.

Solution: Let N_1 be the number of males and N_2 be the number of females employed by the company. Also, let x_1 and x_2 be the monthly average salaries paid to male and female employees and $\bar{x}$ be the mean monthly salary paid to all the employees.

$$\bar{x} = \frac{N_1x_1 + N_2x_2}{N_1 + N_2}$$

or $$500 = \frac{520N_1 + 420N_2}{N_1 + N_2} \quad \text{or} \quad 20N_1 = 80N_2$$

or $$\frac{N_1}{N_2} = \frac{80}{20} = \frac{4}{1}$$

Hence, the males and females are in the ratio of 4 : 1 or 80 per cent are males and 20 per cent are females in those employed by the company.

Short-cut methods for calculating mean

We can simplify the calculations of mean by noticing that if we subtract a constant amount A from each item X to define a new variable $X' = X - A$, the mean $\bar{X}'$ of X' differs from $\bar{X}$ by A. This generally simplifies the calculations and we can then add back the constant A, termed as the *assumed mean:*

$$\bar{X} = A + \bar{X}' = A + \frac{\Sigma f(X')}{\Sigma f}$$

Table 1.4 illustrates the procedure of calculation by short-cut method using the data given in Example 1.13. The choice of A is made in such a manner as to simplify calculation the most, and is generally in the region of the concentration of data.

Table 1.4 *Short-Cut Method of Calculating Mean*

X	(f)	Deviation from Assumed Mean (13) X'	$f(X')$
9	1	–4	–4
10	2	–3	–6
11	3	–2	–6
12	6	–1	–6
			–22
13	10	0	
14	11	+1	+11
15	7	+2	+14
16	3	+3	+9
17	2	+4	+8
18	1	+5	+ 5
			+47
			–22
	$\Sigma f = 46$		$\Sigma fX' = 25$

NOTES

The mean,

$$\overline{X} = A + \frac{\Sigma f(X')}{\Sigma f} = 13 + \frac{25}{46} = 13.54$$

This mean is same as calculated in Example 1.14.

In the case of grouped frequency data, the variable X is replaced by midvalue m, and in the short-cut technique; we subtract a constant value A from each m, so that the formula becomes:

$$\overline{X} = A + \frac{\Sigma f(m - A)}{\Sigma f}$$

In cases where the *class intervals are equal,* we may further simplify calculation by taking the factor i from the variable m–A defining,

$$X' = \frac{m - A}{i}$$

where i is the class width. It can be verified that when X' is defined, then, the mean of the distribution is given by:

$$\overline{X} = A + \frac{\Sigma f(X')}{\Sigma f} \times i$$

The following examples will illustrate the use of the short-cut method.

Example 1.14: The ages of twenty husbands and wives are given in the following table. Form frequency tables showing the relationship between the ages of husbands and wives with class intervals 20–24; 25–29; etc.

Calculate the arithmetic mean of the two groups after the classification.

S.No.	*Age of Husband*	*Age of Wife*
1	28	23
2	37	30
3	42	40
4	25	26
5	29	25
6	47	41
7	37	35
8	35	25
9	23	21
10	41	38
11	27	24
12	39	34
13	23	20
14	33	31
15	36	29
16	32	35
17	22	23
18	29	27
19	38	34
20	48	47

NOTES

Solution:

Calculation of Arithmetic Mean of Husbands' Age

Class Intervals	Midvalues m	Husband Frequency (f_1)	$x_1' = \frac{m-37}{5}$	f_1x_1'
20–24	22	3	–3	–9
25–29	27	5	–2	–10
30–34	32	2	–1	–2
				– 21
35–39	37	6	0	0
40–44	42	2	1	2
45–49	47	2	2	4
				6
		$\Sigma f_1 = 20$		$\Sigma f_1 x_1' = -15$

Husband age, arithmetic mean:

$$\bar{x} = \frac{\Sigma f_1 x_1'}{N} \times i + A = \frac{-15}{20} \times 5 + 37 = 33.25$$

Calculation of Arithmetic Mean of Wives' Age

Class Intervals	Midvalues m	Wife Frequency (f_2)	$x_2' = \frac{m-37}{5}$	f_2x_2'
20–24	22	5	–3	–15
25–29	27	5	–2	–10
30–34	32	4	–1	–4
35–39	37	3	0	0
40–44	42	2	1	2
45–49	47	1	2	2
		$\Sigma f_2 = 20$		$\Sigma f_2 x_2' = -25$

Wife age, arithmetic mean:

$$\bar{x} = \frac{\Sigma f_2 x_2'}{N} \times i + A = \frac{-25}{20} \times 5 + 37 = 30.75$$

Weighted arithmetic mean

In the computation of arithmetic mean we had given equal importance to each observation in the series. This equal importance may be misleading if the individual values constituting the series have different importance as in the following example:

The Raja Toy shop sells

Toy cars at	₹ 3 each
Toy locomotives at	₹ 5 each
Toy aeroplanes at	₹ 7 each
Toy double decker at	₹ 9 each

What shall be the average price of the toys sold, if the shop sells 4 toys, one of each kind?

NOTES

$$\text{Mean price, i.e., } \bar{x} = \frac{\Sigma x}{4} = \text{Rs}\frac{24}{4} = ₹\,6$$

In this case, the importance of each observation (price quotation) is equal in as much as one toy of each variety has been sold. In the above computation of the arithmetic mean, this fact has been taken care of by including 'once only' the price of each toy.

But if the shop sells 100 toys: 50 cars, 25 locomotives, 15 aeroplanes and 10 double deckers, the importance of the four price quotations to the dealer is **not equal** as a source of earning revenue. In fact, their respective importance is equal to the number of units of each toy sold, i.e.,

The importance of toy car	50
The importance of locomotive	25
The importance of aeroplane	15
The importance of double decker	10

It may be noted that 50, 25, 15, 10 are the quantities of the various classes of toys sold. It is for these quantities that the term 'weights' is used in statistical language. Weight is represented by symbol 'w', and Σw represents the sum of weights.

While determining the 'average price of toy sold', these weights are of great importance and are taken into account in the manner illustrated as follows:

$$\bar{x} = \frac{w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4}{w_1 + w_2 + w_3 + w_4} = \frac{\Sigma wx}{\Sigma w}$$

When w_1, w_2, w_3, w_4 are the respective weights of x_1, x_2, x_3, x_4 which in turn represent the price of four varieties of toys, viz., car, locomotive, aeroplane and double decker, respectively.

$$\bar{x} = \frac{(50\times 3)+(25\times 5)+(15\times 7)+(10\times 9)}{50+25+15+10}$$

$$= \frac{(150)+(125)+(105)+(90)}{100} = \frac{470}{100} = ₹\,4.70$$

The table below summarizes the steps taken in the computation of the weighted arithmetic mean.

Table 1.5 Weighted Arithmetic Mean of Toys Sold by the Raja Toy Shop

Toys	*Price per Toy* ₹x	*Number Sold* w	*Price × Weight* xw
Car	3	50	150
Locomotive	5	25	125
Aeroplane	7	15	105
Double Decker	9	10	90
		$\Sigma w = 100$	$\Sigma xw = 470$

$$\Sigma w = 100; \quad \Sigma wx = 470$$

$$\bar{x} = \frac{\sum wx}{\sum w} = \frac{470}{100} = 4.70$$

NOTES

The weighted arithmetic mean is particularly useful where we have to compute the *mean of means.* If we are given two arithmetic means, one for each of two different series, in respect of the *same variable*, and are required to find the arithmetic mean of the combined series, the weighted arithmetic mean is the only suitable method of its determination.

Example 1.15: The arithmetic mean of daily wages of two manufacturing concerns A Ltd. and B Ltd. is ₹ 5 and ₹ 7, respectively. Determine the average daily wages of both concerns if the number of workers employed were 2,000 and 4,000 respectively.

Solution: (*i*) Multiply each average (viz. 5 and 7), by the number of workers in the concern it represents.

(*ii*) Add up the two products obtained in (*i*) above.

(*iii*) Divide the total obtained in (*ii*) by the total number of workers.

Weighted Mean of Mean Wages of A Ltd. and B Ltd.

Manufacturing Concern	*Mean Wages* x	*Workers Employed* w	*Mean Wages × Workers Employed* wx
A Ltd.	5	2,000	10,000
B Ltd.	7	4,000	28,000
		$\sum w = 6,000$	$\sum wx = 38,000$

$$\bar{x} = \frac{\sum wx}{\sum w} = \frac{38,000}{6,000} = ₹\, 6.33$$

The above mentioned examples explain that 'Arithmetic Means and Percentage' are not original data. They are derived figures and their importance is relative to the original data from which they are obtained. This relative importance must be taken into account by weighting while averaging them (means and percentage).

Advantages of mean

- Its concept is familiar to most people and is intuitively clear.
- Every data set has a mean, which is unique and describes the entire data to some degree. For example, when we say that the average salary of a professor is ₹ 25,000 per month, it gives us a reasonable idea about the salaries of professors.
- It is a measure that can be easily calculated.
- It includes all values of the data set in its calculation.
- Its value varies very little from sample to sample taken from the same population.
- It is useful for performing statistical procedures such as computing and comparing the means of several data sets.

Disadvantages of mean

- It is affected by extreme values, and hence, are not very reliable when the data set has extreme values especially when these extreme values are on one side of the ordered data. Thus, a mean of such data is not truly a representative of such data. For example, the average age of three persons of ages 4, 6 and 80 years gives us an average of 30.
- It is tedious to compute for a large data set as every point in the data set is to be used in computations.
- We are unable to compute the mean for a data set that has open-ended classes either at the high or at the low end of the scale.
- The mean cannot be calculated for qualitative characteristics such as beauty or intelligence, unless these can be converted into quantitative figures such as intelligence into IQs.

(ii) Geometric mean

The geometric mean (G) is the nth root of the product of n values.

$$G = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n}$$

The G.M. of 2, 4, 8 is the cube root of their product.

$$G = \sqrt[3]{2 \cdot 4 \cdot 8} = \sqrt[3]{64} = 4$$

If the frequencies of $x_1, x_2, \ldots\ldots, x_k$ are respectively $f_1, f_2, \ldots\ldots, f_k (\Sigma f = n)$

$$G = \sqrt[n]{x_1^{f_1} \cdot x_z^{f_2} \ldots\ldots x_k^{f_k}}$$

Logarithms may be used in the calculation of G.M.

$$\text{Log } G = \frac{1}{n}[f_1 \log x_1 + f_2 \log x_2 + \ldots + f_k \log x_k] = \frac{\Sigma f \log x}{n}$$

$$G = \text{Antilog } \frac{1}{n} \Sigma f \log x$$

If there are no frequencies, $G = (x_1 x_2 \ldots x_n)^{\frac{1}{n}}$ and $\log G = \frac{1}{n} \Sigma \log x$

Merits and uses of geometric mean

Most of the properties and merits of G.M. resemble those of A.M.

- The GM takes into account all the items in the data and condenses them into one respectative value.
- It has a downward bias. It gives more weight to smaller values than to larger values.
- It is determinate. For the same data there cannot be two geometric means.
- It balances the ratios of the values on either side of the data. It is ideally suited to average rates of change such as index numbers and ratios between measures and percentages.
- It is amenable to algebraic manipulations like the A.M.

Demerits of geometric mean

- It is difficult to use and to compute.
- It is determined for positive values and cannot be used for negative values of zero. A zero will convert the whole product into zero.

(iii) Median

The second measure of central tendency that has a wide usage in statistical works is the median. Median is that *value* of a variable which divides the series in such a manner that the number of items below it is equal to the number of items above it. Half the total number of observations lie below the median, and half above it. The median is thus a positional average.

The median of ungrouped data is found easily if the items are first arranged in order of the magnitude. The median may then be located simply by counting, and its value can be obtained by reading the value of the middle observations. If we have five observations whose values are 8, 10, 1, 3 and 5, the values are first arrayed: 1, 3, 5, 8 and 10. It is now apparent that the value of the median is 5, since two observations are below that value and two observations are above it. When there is an even number of cases, there is no actual middle item and the median is taken to be the average of the values of the items lying on either side of $(N+1)/2$, where N is the total number of items. Thus, if the values of six items of a series are 1, 2, 3, 5, 8 and 10, then the median is the value of item number $(6+1)/2 = 3.5$, which is approximated as the average of the third and the fourth items, i.e., $(3+5)/2 = 4$.

Thus, the steps required for obtaining median are:

1. Arrange the data as an array of increasing magnitude.
2. Obtain the value of the $(N+1)/2$th item.

Even in the case of grouped data, the procedure for obtaining median is straightforward as long as the variable is discrete or non-continuous as is clear from the following example.

Example 1.16: Obtain the median size of shoes sold from the following data.

Number of Shoes Sold by Size in One Year

Size	*Number of Pairs*	*Cumulative Total*
5	30	30
$5\frac{1}{2}$	40	70
6	50	120
$6\frac{1}{2}$	150	270
7	300	570
$7\frac{1}{2}$	600	1170
8	950	2120
$8\frac{1}{2}$	820	2940
9	750	3690
$9\frac{1}{2}$	440	4130
10	250	4380
$10\frac{1}{2}$	150	4530
11	40	4570
$11\frac{1}{2}$	39	4609
	Total	4609

Solution: Median, is the value of $\frac{(N+1)}{2}$th $= \frac{4609+1}{2}$th = 2305th item. Since the items are already arranged in ascending order (size-wise), the size of 2305th item is easily determined by constructing the cumulative frequency. Thus, the median size of shoes sold is 8½, the size of 2305th item.

In the case of grouped data with continuous variable, the determination of median is a bit more involved. Consider the following table where the data relating to the distribution of male workers by average monthly earnings is given. Clearly the median of 6291 is the earnings of (6291 + 1)/2 = 3146th worker arranged in ascending order of earnings.

From the cumulative frequency, it is clear that this worker has his income in the class interval 67.5–72.5. But, it is impossible to determine his exact income. We therefore, resort to approximation by assuming that the 795 workers of this class are distributed *uniformly* across the interval 67.5 to 72.5. The median worker is (3146–2713) = 433rd of these 795, and hence, the value corresponding to him can be approximated as,

$$67.5 + \frac{433}{795} \times (72.5 - 67.5) = 67.5 + 2.73 = 70.23$$

Distribution of Male Workers by Average Monthly Earnings

Group No.	*Monthly Earnings (₹)*	*No. of Workers*	*Cumulative No. of Workers*
1	27.5–32.5	120	120
2	32.5–37.5	152	272
3	37.5–42.5	170	442
4	42.5–47.5	214	656
5	47.5–52.5	410	1066
6	52.5–57.5	429	1495
7	57.5–62.5	568	2063
8	62.5–67.5	650	2713
9	67.5–72.5	795	3508
10	72.5–77.5	915	4423
11	77.5–82.5	745	5168
12	82.5–87.5	530	5698
13	87.5–92.5	259	5957
14	92.5–97.5	152	6109
15	97.5–102.5	107	6216
16	102.5–107.5	50	6266
17	107.5–112.5	25	6291
			Total 6291

The value of the median can thus be put in the form of the formula,

$$Me = l + \frac{\frac{N+1}{2} - C}{f} \times i$$

Where l is the lower limit of the median class, i its width, f its frequency, C the cumulative frequency upto (but not including) the median class, and N is the total number of cases.

Finding median by graphical analysis

The median can quite conveniently be determined by reference to the ogive which plots the cumulative frequency against the variable. The value of the item below which half the items lie, can easily be read from the ogive.

Example 1.17: Obtain the median of data given in the following table.

Monthly Earnings	*Frequency*	*Less Than*	*More Than*
27.5	—	0	6291
32.5	120	120	6171
37.5	152	272	6019
42.5	170	442	5849
47.5	214	656	5635
52.5	410	1066	5225
57.5	429	1495	4796
62.5	568	2063	4228
67.5	650	2713	3578
72.5	795	3508	2783
77.5	915	4423	1868
82.5	745	5168	1123
87.5	530	5698	593
92.5	259	5957	334
97.5	152	6109	182
102.5	107	6216	75
107.5	50	6266	25
112.5	25	6291	0

Solution: It is clear that this is grouped data. The first class is 27.5–32.5, whose frequency is 120, and the last class is 107.5–112.5, whose frequency is 25. Figure 1.6 shows the ogive of less than cumulative frequency. The median is the value below which $N/2$ items lie, is $6291/2 = 3145.5$ items lie, which is read of from Figure 1.7 as about 70. More accuracy than this is unobtainable because of the space limitation on the earning scale.

NOTES

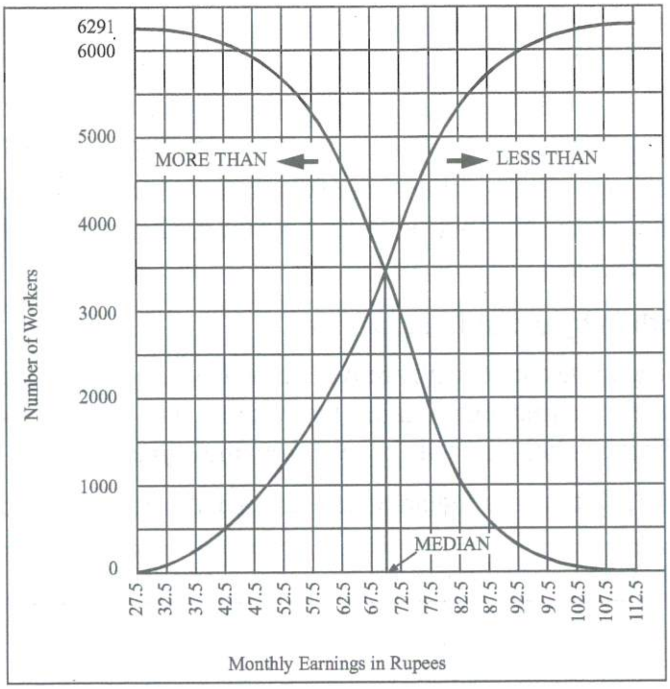


Fig. 1.6 Median Determination by Plotting Less than and More than Cumulative Frequency

The median can also be determined by plotting both 'less than' and 'more than' cumulative frequency as shown in Figure 1.6. It should be obvious that the two curves should intersect at the median of the data.

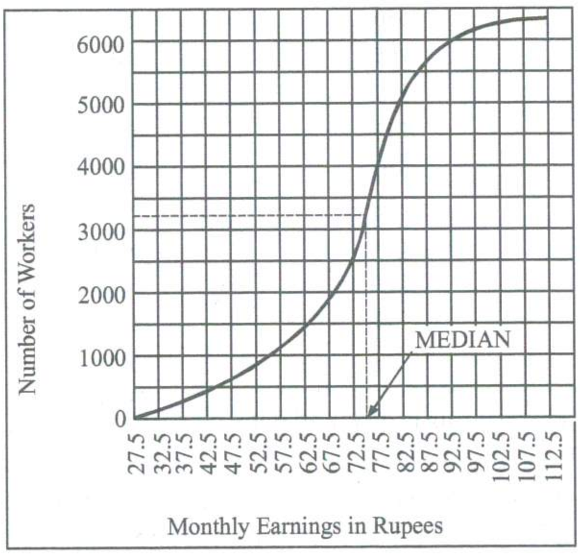


Fig. 1.7 Median

Advantages of median

- Median is a positional average and hence the extreme values in the data set do not affect it as much as they do to the mean.
- Median is easy to understand and can be calculated from any kind of data, even from grouped data with open-ended classes.

- We can find the median even when our data set is qualitative and can be arranged in the ascending or the descending order, such as average beauty or average intelligence.
- Similar to mean, median is also unique, meaning that, there is only one median in a given set of data.
- Median can be located visually when the data is in the form of ordered data.
- The sum of absolute differences of all values in the data set from the median value is minimum. This means that, it is less than any other value of central tendency in the data set, which makes it more central in certain situations.

Disadvantages of median

- The data must be arranged in order to find the median. This can be very time consuming for a large number of elements in the data set.
- The value of the median is affected more by sampling variations. Different samples from the same population may give significantly different values of the median.
- The calculation of median in case of grouped data is based on the assumption that the values of observations are evenly spaced over the entire class interval and this is usually not so.
- Median is comparatively less stable than mean, particularly for small samples, due to fluctuations in sampling.
- Median is not suitable for further mathematical treatment. For example, we cannot compute the median of the combined group from the median values of different groups.

(iv) Mode

The mode is that value of the variable which occurs or repeats itself the greatest number of times. The mode is the most 'fashionable' size in the sense that it is the most common and typical, and is defined by Zizek as 'the value occurring most frequently in a series (or group of items) and around which the other items are distributed most densely'.

The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It is the most frequent or the most common value, provided that a sufficiently large number of items are available, to give a smooth distribution. It will correspond to the value of the maximum point (ordinate), of a frequency distribution if it is an 'ideal' or smooth distribution. It may be regarded as the most typical of a series of values. The modal wage, for example, is the wage received by more individuals than any other wage. The modal 'hat' size is that, which is worn by more persons than any other single size.

It may be noted that the occurrence of one or a few extremely high or low values has no effect upon the mode. If a series of data are unclassified, not have been either arrayed or put into a frequency distribution, the mode cannot be readily located.

Taking first an extremely simple example, if seven men are receiving daily wages of ₹ 5, 6, 7, 7, 7, 8 and 10, it is clear that the modal wage is ₹ 7 per day. If

NOTES

we have a series such as 2, 3, 5, 6, 7, 10 and 11, it is apparent that there is no mode.

There are several methods of estimating the value of the mode. But, it is seldom that the different methods of ascertaining the mode give us identical results. Consequently, it becomes necessary to decide as to which method would be most suitable for the purpose in hand. In order that a choice of the method may be made, we should understand each of the methods and the differences that exist among them.

The four important methods of estimating mode of a series are: (*i*) Locating the most frequently repeated value in the array; (*ii*) Estimating the mode by interpolation; (*iii*) Locating the mode by graphic method; and (*iv*) Estimating the mode from the mean and the median. Only the last three methods are discussed in this unit.

Estimating the mode by interpolation

In the case of continuous frequency distributions, the problem of determining the value of the mode is not so simple as it might have appeared from the foregoing description. Having located the modal class of the data, the next problem in the case of continuous series is to interpolate the value of the mode within this 'modal' class.

The interpolation is made by the use of any one of the following formulae:

$$(i)\ Mo = l_1 + \frac{f_2}{f_0 + f_2} \times i; \quad (ii)\ Mo = l_2 - \frac{f_0}{f_0 + f_2} \times i$$

$$(iii)\ Mo = l_1 + \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \times i$$

Where l_1 is the lower limit of the modal class, l_2 is the upper limit of the modal class, f_0 equals the frequency of the preceding class in value, f_1 equals the frequency of the modal class in value, f_2 equals the frequency of the following class (class next to modal class) in value, and i equals the interval of the modal class.

Example 1.18: Determine the mode for the data given in the following table.

Wage Group	*Frequency (f)*
14 — 18	6
18 — 22	18
22 — 26	19
26 — 30	12
30 — 34	5
34 — 38	4
38 — 42	3
42 — 46	2
46 — 50	1
50 — 54	0
54 — 58	1

Solution: In the given data, 22 – 26 is the modal class since it has the largest frequency. The lower limit of the modal class is 22, its upper limit is 26, its frequency is 19, the frequency of the preceding class is 18, and of the following class is 12. The class interval is 4. Using the various methods of determining mode, we have,

$$(i)\ Mo = 22 + \frac{12}{18+12} \times 4 \qquad (ii)\ Mo = 26 - \frac{18}{18+12} \times 4$$

$$= 22 + \frac{8}{5} \qquad\qquad = 26 - \frac{12}{5}$$

$$= 23.6 \qquad\qquad = 23.6$$

$$(iii)\ Mo = 22 + \frac{19-18}{(19-18)+(19-12)} \times 4 = 22 + \frac{4}{8} = 22.5$$

NOTES

In formulae (*i*) and (*ii*), the frequency of the classes adjoining the modal class is used to pull the estimate of the mode away from the midpoint towards either the upper or lower class limit. In this particular case, the frequency of the class preceding the modal class is more than the frequency of the class following and therefore, the estimated mode is less than the midvalue of the modal class. This seems quite logical. If the frequencies are more on one side of the modal class than on the other it can be reasonably concluded that the items in the modal class are concentrated more towards the class limit of the adjoining class with the larger frequency.

The formula (*iii*) is also based on a logic similar to that of (*i*) and (*ii*). In this case, to interpolate the value of the mode within the modal class, the differences between the frequency of the modal class, and the respective frequencies of the classes adjoining it are used. This formula usually gives results better than the values obtained by the other and exactly equal to the results obtained by graphic method. The formulae (*i*) and (*ii*) give values which are different from the value obtained by formula (*iii*) and are more close to the central point of modal class. If the frequencies of the class adjoining the modal are equal, the mode is expected to be located at the midvalue of the modal class, but if the frequency on one of the sides is greater, the mode will be pulled away from the central point. It will be pulled more and more if the difference between the frequencies of the classes adjoining the modal class is higher and higher. In Example 1.18, the frequency of the modal class is 19 and that of preceding class is 18. So, the mode should be quite close to the lower limit of the modal class. The midpoint of the modal class is 24 and lower limit of the modal class is 22.

Locating the mode by the graphic method

The upper corners of the rectangle over the modal class have been joined by straight lines to those of the adjoining rectangles as shown in the diagram; the right corner to the corresponding one of the adjoining rectangle on the left, etc. If a perpendicular is drawn from the point of intersection of these lines, we have a value for the mode indicated on the base line. The graphic approach is, in principle, similar to the arithmetic interpolation explained earlier.

The mode may also be determined graphically from an ogive or cumulative frequency curve. It is found by drawing a perpendicular to the base from that point on the curve where the curve is most nearly vertical, i.e., steepest (in other words, where it passes through the greatest distance vertically and smallest distance horizontal). The point where it cuts the base gives us the value of the mode. How accurately this method determines the mode is governed by: (*i*) The shape of the ogive, (*ii*) The scale on which the curve is drawn.

NOTES

Estimating the mode from the mean and the median

There usually exists a relationship among the mean, median and mode for moderately asymmetrical distributions. If the distribution is symmetrical, the mean, median and mode will have identical values, but if the distribution is skewed (moderately) the mean, median and mode will pull apart. If the distribution tails off towards higher values, the mean and the median will be greater than the mode. If it tails off towards lower values, the mode will be greater than either of the other two measures. In either case, the median will be about one-third as far away from the mean as the mode is. This means that,

$$\text{Mode} = \text{Mean} - 3\,(\text{Mean} - \text{Median})$$
$$= 3\,\text{Median} - 2\,\text{Mean}$$

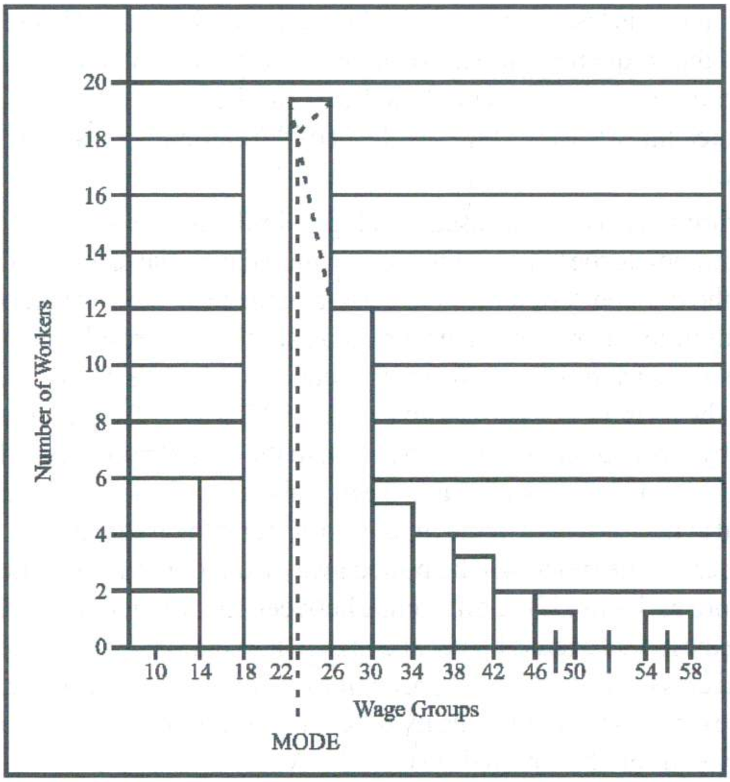


Fig. 1.8 Method of Mode Determination by Graphic Interpolation

In the case of the average monthly earnings, the mean is 68.53 and the median is 70.2. If these values are substituted in the above formula, we get,

$$\text{Mode} = 68.5 - 3(68.5 - 70.2)$$
$$= 68.5 + 5.1 = 73.6$$

According to the formula used earlier,

$$\text{Mode} = l_1 + \frac{f_2}{f_0 + f_2} \times i$$
$$= 72.5 + \frac{745}{795 + 745} \times 5$$
$$= 72.5 + 2.4 = 74.9$$

OR

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

$$= 72.5 + \frac{915 - 795}{2 \times 915 - 795 - 745} \times 5$$

$$= 72.5 + \frac{120}{290} \times 5 = 74.57$$

The difference between the two estimates is due to the fact that the assumption of relationship between the mean, median and mode may not always be true which is obviously not valid in this case.

Example 1.19: (*i*) In a moderately symmetrical distribution, the mode and mean are 32.1 and 35.4 respectively. Calculate the median.

(*ii*) If the mode and median of moderately asymmetrical series are respectively 16" and 15.7", what would be its most probable median?

(*iii*) In a moderately skewed distribution, the mean and the median are respectively 25.6 and 26.1 inches. What is the mode of the distribution?

Solution: (*i*) We know,

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})$$

or $$3\text{ Median} = \text{Mode} + 2\text{ Mean}$$

or $$\text{Median} = \frac{32.1 + 2 \times 35.4}{3}$$

$$= \frac{102.9}{3}$$

$$= 34.3$$

(*ii*) $$2\text{ Mean} = 3\text{ Median} - \text{Mode}$$

or $$\text{Mean} = \frac{1}{2}(3 \times 15.7 - 16.0) = \frac{31.1}{2} = 15.55$$

(*iii*) $$\text{Mode} = 3\text{ Median} - 2\text{ Mean}$$

$$= 3 \times 26.1 - 2 \times 25.6 = 78.3 - 51.2 = 27.1$$

Advantages of mode

- Similar to median, the mode is not affected by extreme values in the data.
- Its value can be obtained in open-ended distributions without ascertaining the class limits.
- It can be easily used to describe qualitative phenomenon. For example, if most people prefer a certain brand of tea, then this will become the modal point.
- Mode is easy to calculate and understand. In some cases, it can be located simply by observation or inspection.

Disadvantages of mode

- Quite often, there is no modal value.
- It can be bi-modal or multi-modal, or it can have all modal values making its significance more difficult to measure.
- If there is more than one modal value, the data is difficult to interpret.
- A mode is not suitable for algebraic manipulations.
- Since the mode is the value of maximum frequency in the data set, it cannot be rigidly defined if such frequency occurs at the beginning or at the end of the distribution.
- It does not include all observations in the data set, and hence, less reliable in most of the situations.

(v) Quartiles, Deciles and Percentiles

Some measures other than measures of central tendency are often employed when summarizing or describing a set of data where it is necessary to divide the data into equal parts. These are positional measures and are called quantiles and consist of quartiles, deciles and percentiles. The quartiles divide the data into four equal parts. The deciles divide the total ordered data into ten equal parts and percentiles divide the data into 100 equal parts. Consequently, there are three quartiles, nine deciles and 99 percentiles. The quartiles are denoted by the symbol Q so that Q_1 will be such point in the ordered data which has 25 per cent of the data below and 75 per cent of the data above it. In other words Q_1 is the value corresponding to $\left(\frac{n+1}{4}\right)$th ordered observation. Similarly, Q_2 divides the data in the middle, and is also equal to the median and its value Q_2 is given by:

$$Q_2 = \text{The value of } 2\left(\frac{n+1}{4}\right)\text{th ordered observation in the data.}$$

Similarly, we can calculate the values of various deciles. For instance,

$$D_1 = \left(\frac{n+1}{10}\right)\text{th observaton in the data, and}$$

$$D_7 = 7\left(\frac{n+1}{10}\right)\text{th observation in the ordered data.}$$

Percentiles are generally used in the research area of education where people are given standard tests and it is desirable to compare the relative position of the subject's performance on the test. Percentiles are similarly calculated as:

$$P_7 = 7\left(\frac{n+1}{100}\right)\text{th observation in the ordered data.}$$

and,

$$P_{69} = 69\left(\frac{n+1}{100}\right)\text{th observation in the ordered data.}$$

Quartiles

The formula for calculating the values of quartiles for grouped data is given as follows.

$$Q = L + (j/f)C$$

where,

Q = The quartile under consideration.

L = Lower limit of the class interval which contains the value of Q.

j = The number of units we lack from the class interval which contains the value of Q, in reaching the value of Q.

f = Frequency of the class interval containing Q.

C = Size of the class interval.

Let us assume we took the data of the ages of 100 students and a frequency distribution for this data has been constructed as shown.

The frequency distribution is as follows:

Ages (CI)	*Mid-point (X)*	*(f)*	*f(X)*	*f(X)*^{2}
16 and upto 17	16.5	4	66	1089.0
17 and upto 18	17.5	14	245	4287.5
18 and upto 19	18.5	18	333	6160.5
19 and upto 20	19.5	28	546	10647.0
20 and upto 21	20.5	20	410	8405.0
21 and upto 22	21.5	12	258	5547.0
22 and upto 23	22.5	4	90	2025.0
		Totals = 100	1948	38161

In our case, in order to find Q_1, where Q_1 is the cut off point so that 25 per cent of the data is below this point and 75 per cent of the data is above, we see that the first group has 4 students and the second group has 14 students making a total of 18 students. Since Q_1 cuts off at 25 students, it is the third class interval which contains Q_1. This means that the value of L in our formula is 18.

Since we already have 18 students in the first two groups, we need 7 more students from the third group to make it a total of 25 students, which is the value of Q_1. Hence, the value of (j) is 7. Also, since the frequency of this third class interval which contains Q_1 is 18, the value of (f) in our formula is 18. The size of the class interval C is given as 1. Substituting these values in the formula for Q, we get

$$Q_1 = 18 + (7/18)1$$
$$= 18 + .38 = 18.38$$

This means that 25 per cent of the students are below 18.38 years of age and 75 per cent are above this age.

Similarly, we can calculate the value of Q_2, using the same formula. Hence,

$$Q_2 = L + (j/f)C$$
$$= 19 + (14/28)1 = 19.5$$

This also happens to be the median.

NOTES

By using the same formula and same logic we can calculate the values of all deciles as well as percentiles.

We have defined the median as the value of the item which is located at the centre of the array. We can define other measures which are located at other specified points. Thus, the *N*th percentile of an array is the value of the item such that *N* per cent items lie below it. Clearly then the N_{th} percentile P_n of grouped data is given by

$$P_n = l + \frac{\frac{nN}{100} - C}{f} \times i$$

where *l* is the lower limit of the class in which *nN*/100th item lies, *i* its width, *f* its frequency, *C* the cumulative frequency upto (but not including) this class, and *N* is the total number of items.

We similarly define the *N*th decile as the value of the item below which (*nN*/10) items of the array lie. Clearly,

$$D_n = P_{10n} = l + \frac{\frac{nN}{10} - C}{f} \times i$$

where the symbols have the obvious meanings.

The other most commonly referred to measures of location are the quartiles. Thus, *n*th quartile is the value of the item which lies at the *n*(*N*/5)th item. Clearly Q_2, the second quartile is the median, for grouped data.

$$Q_n = P_{25n} = l + \frac{\frac{nN}{4} - C}{f} \times i$$

(vi) Measures of Dispersion

A measure of dispersion, or simply dispersion may be defined as statistics signifying the extent of the scatteredness of items around a measure of central tendency.

A measure of dispersion may be expressed in an 'absolute form', or in a 'relative form'. It is said to be in an absolute form when it states the actual amount by which the value of an item on an average deviates from a measure of central tendency. Absolute measures are expressed in concrete units, i.e., units in terms of which the data have been expressed, e.g., rupees, centimetres, kilograms, etc., and are used to describe frequency distribution.

A relative measure of dispersion computed is a quotient by dividing the absolute measures by a quantity in respect to which absolute deviation has been computed. It is as such a pure number and is usually expressed in a percentage form. Relative measures are used for making comparisons between two or more distributions.

A measure of dispersion should possess all those characteristics which are considered essential for a measure of central tendency, viz.

- It should be based on all observations.
- It should be readily comprehensible.
- It should be fairly easily calculated.
- It should be affected as little as possible by fluctuations of sampling.
- It should be amenable to algebraic treatment.

The following are some common measures of dispersion:

(*i*) The range, (*ii*) the semi-interquartile range or the quartile deviation, (*iii*) the mean deviation, and (*iv*) the standard deviation. Of these, the standard deviation is the best measure. We describe these measures in the following sections.

(a) Range

The crudest measure of dispersion is the range of the distribution. The range of any series is the difference between the highest and the lowest values in the series. If the marks received in an examination taken by 248 students are arranged in ascending order, then the range will be equal to the difference between the highest and the lowest marks.

In a frequency distribution, the range is taken to be the difference between the lower limit of the class at the lower extreme of the distribution and the upper limit of the class at the upper extreme.

Table 1.6 *Weekly Earnings of Labourers in Four Workshops of the Same Type*

Weekly earnings	*No. of workers*			
	Workshop A	*Workshop B*	*Workshop C*	*Workshop D*
15–16	...	...	2	...
17–18	...	2	4	...
19–20	...	4	4	4
21–22	10	10	10	14
23–24	22	14	16	16
25–26	20	18	14	16
27–28	14	16	12	12
29–30	14	10	6	12
31–32	...	6	6	4
33–34	...	...	2	2
35–36	...	...	...	...
37–38	...	...	4	...
Total	80	80	80	80
Mean	25.5	25.5	25.5	25.5

Consider the data on weekly earning of worker on four workshops given in the Table 1.6. We note the following:

Workshop	*Range*
A	9
B	15
C	23
D	15

NOTES

From these figures, it is clear that the greater the range, the greater is the variation of the values in the group.

The range is a measure of absolute dispersion and as such cannot be usefully employed for comparing the variability of two distributions expressed in different units. The amount of dispersion measured, say, in pounds, is not comparable with dispersion measured in inches. So the need of measuring relative dispersion arises.

An absolute measure can be converted into a relative measure if we divide it by some other value regarded as standard for the purpose. We may use the mean of the distribution or any other positional average as the standard.

For Table 1.6, the relative dispersion would be:

$$\text{Workshop } A = \frac{9}{25.5} \qquad \text{Workshop } C = \frac{23}{25.5}$$

$$\text{Workshop } B = \frac{15}{25.5} \qquad \text{Workshop } D = \frac{15}{25.5}$$

An alternate method of converting an absolute variation into a relative one would be to use the total of the extremes as the standard. This will be equal to dividing the difference of the extreme items by the total of the extreme items. Thus,

$$\text{Relative Dispersion} = \frac{\text{Difference of extreme items, i.e, Range}}{\text{Sum of extreme items}}$$

The relative dispersion of the series is called the coefficient or ratio of dispersion. In our example of weekly earnings of workers considered earlier, the coefficients would be:

$$\text{Workshop } A = \frac{9}{21+30} = \frac{9}{51} \qquad \text{Workshop } B = \frac{15}{17+32} = \frac{15}{49}$$

$$\text{Workshop } C = \frac{23}{15+38} = \frac{23}{53} \qquad \text{Workshop } D = \frac{15}{19+34} = \frac{15}{53}$$

Merits and limitations of range

Merits

Of the various characteristics that a good measure of dispersion should possess, the range has only two, viz (i) it is easy to understand, and (ii) its computation is simple.

Limitations

Besides the aforesaid two qualities, the range does not satisfy the other test of a good measure and hence it is often termed as a crude measure of dispersion.

The following are the limitations that are inherent in the range as a concept of variability:

(i) Since it is based upon two extreme cases in the entire distribution, the range may be considerably changed if either of the extreme cases happens to drop out, while the removal of any other case would not affect it at all.

(ii) It does not tell anything about the distribution of values in the series relative to a measure of central tendency.

(iii) It cannot be computed when distribution has open-end classes.

(iv) It does not take into account the entire data. These can be illustrated by the following illustration. Consider the data given in Table 1.7.

Table 1.7 *Distribution with the Same Number of Cases, but Different Variability*

Class	*No. of students*		
	Section A	*Section B*	*Section C*
0–10	...	...	...
10–20	1	...	...
20–30	12	12	19
30–40	17	20	18
40–50	29	35	16
50–60	18	25	18
60–70	16	10	18
70–80	6	8	21
80–90	11	...	...
90–100	...	...	...
Total	110	110	110
Range	80	60	60

The table is designed to illustrate three distributions with the same number of cases but different variability. The removal of two extreme students from section *A* would make its range equal to that of *B* or *C*.

The greater range of *A* is not a description of the entire group of 110 students, but of the two most extreme students only. Further, though sections *B* and *C* have the same range, the students in section B cluster more closely around the central tendency of the group than they do in section *C*. Thus, the range fails to reveal the greater homogeneity of *B* or the greater dispersion of *C*. Due to this defect, it is seldom used as a measure of dispersion.

Specific uses of range

In spite of the numerous limitations of the range as a measure of dispersion, there are the following circumstances when it is the most appropriate one:

(i) In situations where the extremes involve some hazard for which preparation should be made, it may be more important to know the most extreme cases to be encountered than to know anything else about the distribution. For example, an explorer, would like to know the lowest and the highest temperatures on record in the region he is about to enter; or an engineer would like to know the maximum rainfall during 24 hours for the construction of a storem water drain.

(ii) In the study of prices of securities, range has a special field of activity. Thus to highlight fluctuations in the prices of shares or bullion it is a common practice to indicate the range over which the prices have moved during a

NOTES

certain period of time. This information, besides being of use to the operators, gives an indication of the stability of the bullion market, or that of the investment climate.

(iii) In statistical quality control the range is used as a measure of variation. We, e.g., determine the range over which variations in quality are due to random causes, which is made the basis for the fixation of control limits.

(b) Quartile Deviation

Another measure of dispersion, much better than the range, is the semi-interquartile range, usually termed as 'quartile deviation'. As stated in the previous unit, quartiles are the points which divide the array in four equal parts. More precisely, Q_1 gives the value of the item 1/4th the way up the distribution and Q_3 the value of the item 3/4th the way up the distribution. Between Q_1 and Q_3 are included half the total number of items. The difference between Q_1 and Q_3 includes only the central items but excludes the extremes. Since under most circumstances, the central half of the series tends to be fairly typical of all the items, the interquartile range ($Q_3 - Q_1$) affords a convenient and often a good indicator of the absolute variability. The larger the interquartile range, the larger the variability.

Usually, one-half of the difference between Q_3 and Q_1 is used and to it is given the name of quartile deviation or semi-interquartile range. The interquartile range is divided by two for the reason that half of the interquartile range will, in a normal distribution, be equal to the difference between the median and any quartile. This means that 50 per cent items of a normal distribution will lie within the interval defined by the median plus and minus the semi-interquartile range.

Symbolically:

$$Q.D. = \frac{Q_3 - Q_1}{2} \qquad ...(1)$$

Let us find quartile deviations for the weekly earnings of labour in the four workshop whose data is given in Table 1.6. The computations are as shown in Table 1.8.

As shown in the table, Q.D. of workshop A is ₹ 2.12 and median value in 25.3. This means that if the distribution is symmetrical the number of workers whose wages vary between (25.3–2.1) = ₹ 23.2 and (25.3 + 2.1) = ₹ 27.4, shall be just half of the total cases. The other half of the workers will be more than ₹ 2.1 removed from the median wage. As this distribution is not symmetrical, the distance between Q_1 and the median Q_2 is not the same as between Q_3 and the median. Hence the interval defined by median plus and minus semi inter-quartile range will not be exactly the same as given by the value of the two quartiles. Under such conditions the range between ₹ 23.2 and ₹ 27.4 will not include precisely 50 per cent of the workers.

If quartile deviation is to be used for comparing the variability of any two series, it is necessary to convert the absolute measure to a coefficient of quartile deviation. To do this the absolute measure is divided by the average size of the two quartile.

Symbolically:

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad ...(2)$$

Applying this to our illustration of four workshops, the coefficients of Q.D. are as given below.

Table 1.8 *Calculation of Quartile Deviation*

		Workshop A	*Workshop B*	*Workshop C*	*Workshop D*
Location of Q_2	$\frac{N}{2}$	$\frac{80}{2}=40$	$\frac{80}{2}=40$	$\frac{80}{2}=40$	$\frac{80}{2}=40$
	Q_2	$24.5+\frac{40-30}{22}\times 2$ $=24.5+0.9$ $=25.4$	$24.5+\frac{40-30}{18}\times 2$ $=24.5+1.1$ $=25.61$	$24.5+\frac{40-30}{16}\times 2$ $=24.5+0.75$ $=25.25$	$24.5+\frac{40-30}{16}\times 2$ $=24.5+0.75$ $=25.25$
Location of Q_1	$\frac{N}{4}$	$\frac{80}{4}=20$	$\frac{80}{4}=20$	$\frac{80}{4}=20$	$\frac{80}{4}=20$
	Q_1	$22.5+\frac{20-10}{22}\times 2$ $=22.5+.91$ $=23.41$	$22.5+\frac{20-16}{14}\times 2$ $=22.5+.57$ $=23.07$	$20.5+\frac{20-10}{10}\times 2$ $=20.5+2$ $=22.5$	$22.5+\frac{20-18}{16}\times 2$ $=22.5+.25$ $=22.75$
Location of Q_3	$\frac{3N}{4}$	$3\times\frac{80}{4}=60$	60	60	60
	Q_3	$26.5+\frac{60-52}{14}\times 2$	$26.5+\frac{60-48}{16}\times 2$	$26.5+\frac{60-50}{12}\times 2$	
$26.5+\frac{60-50}{12}\times 2$					
		$=26.5+1.14$ $=27.64$	$=26.5+1.5$ $=28.0$	$=26.5+1.67$ $=28.17$	$=26.5+1.67$ $=28.17$
Quartile Deviation	$\frac{Q_3-Q_1}{2}$	$\frac{27.64-23.41}{2}$ $=\frac{4.23}{2}=$ ₹ 2.12	$\frac{28-23.07}{2}$ $=\frac{4.93}{2}=$ ₹ 2.46	$\frac{28.17-22.5}{2}$ $=\frac{5.67}{2}=$ ₹ 2.83	$\frac{28.17-22.75}{2}$ $=\frac{5.42}{2}=$ ₹. 2.71
Coefficient of quartile deviation	$\frac{Q_3-Q_1}{Q_3+Q_1}=$	$\frac{27.64-23.41}{27.64+23.41}$ $=0.083$	$\frac{28-23.07}{28+23.07}$ $=0.097$	$\frac{28.17-22.5}{28.17+22.5}$ $=0.112$	$\frac{28.17-22.75}{28.17+22.75}$ $=0.106$

Characteristics of quartile deviation

- The size of the quartile deviation gives an indication about the uniformity or otherwise of the size of the items of a distribution. If the quartile deviation is small it denotes large uniformity. Thus, a coefficient of quartile deviation may be used for comparing uniformity or variation in different distributions.
- Quartile deviation is not a measure of dispersion in the sense that it does not show the scatter around an average, but only a distance on scale. Consequently, quartile deviation is regarded as a measure of partition.
- It can be computed when the distribution has open-end classes.

Limitations of quartile deviation

Except for the fact that its computation is simple and it is easy to understand, a quartile deviation does not satisfy any other test of a good measure of variation.

(c) Mean Deviation

A weakness of the measures of dispersion discussed earlier, based upon the range or a portion thereof, is that the precise size of most of the variants has no effect on the result. As an illustration, the quartile deviation will be the same whether the variates between Q_1 and Q_3 are concentrated just above Q_1 or they are spread uniformly from Q_1 to Q_3. This is an important defect from the viewpoint of measuring the divergence of the distribution from its typical value. The mean deviation is employed to answer the objection.

Mean deviation also called average deviation, of a frequency distribution is the mean of the absolute values of the deviation from some measure of central tendency. In other words, mean deviation is the arithmetic average of the variations (deviations) of the individual items of the series from a measure of their central tendency.

We can measure the deviations from any measure of central tendency, but the most commonly employed ones are the median and the mean. The median is preferred because it has the important property that the average deviation from it is the least.

Calculation of the mean deviation then involves the following steps:

(*a*) Calculate the median (or the mean) *Me* (or $\overline{X}$).

(*b*) Record the deviations $|d| = |x - Me|$ of each of the items, ignoring the sign.

(*c*) Find the average value of deviations.

$$\text{Mean Deviation} = \frac{\Sigma|d|}{N} \qquad ...(3)$$

Example 1.20: Calculate the mean deviation from the following data giving marks obtained by 11 students in a class test.

14, 15, 23, 20, 10, 30, 19, 18, 16, 25, 12.

Solution: $\text{Median} = \text{Size of } \frac{11+1}{2} \text{ th item}$

$= \text{size of 6th item} = 18.$

Serial No.	*Marks*	$\lvert x - Median \rvert$ $\lvert d \rvert$
1	10	8
2	12	6
3	14	4
4	15	3
5	16	2

6	18	0
7	19	1
8	20	2
9	23	5
10	25	7
11	30	12
		$\Sigma\|d\| = 50$

$$\text{Mean deviation from median} = \frac{\Sigma |d|}{N}$$

$$= \frac{50}{11} = 4.5 \text{ marks.}$$

For grouped data, it is easy to see that the mean deviation is given by

$$\text{Mean deviation, M.D.} = \frac{\Sigma f |d|}{\Sigma f} \qquad ...(4)$$

where $|d| = |x - \text{median}|$ for grouped discrete data, and $|d| = M - \text{median}|$ for grouped continuous data with M as the mid-value of a particular group. The following examples illustrate the use of this formula.

Example 1.21: Calculate the mean deviation from the following data

Size of item	6	7	8	9	10	11	12
Frequency	3	6	9	13	8	5	4

Solution:

Size	*Frequency* f	*Cumulative frequency*	*Deviations from median (9)* $\|d\|$	$f\|d\|$
6	3	3	3	9
7	6	9	2	12
8	9	18	1	9
9	13	31	0	0
10	8	39	1	8
11	5	44	2	10
12	4	48	3	12
	48			60

$$\text{Median} = \text{the size of } \frac{48+1}{2} = 24.5\text{th item which is 9.}$$

Therefore, deviations d are calculated from 9, i.e., $|d| = |x - 9|$.

$$\text{Mean deviation} = \frac{\Sigma f |d|}{\Sigma f} = \frac{60}{48} = 1.25$$

Example 1.22: Calculate the mean deviation from the following data:

x	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
f	18	16	15	12	10	5	2	2

NOTES

Solution:

This is a frequency distribution with continuous variable. Thus, deviations are calculated from mid-values.

x	*Mid-value*	*f*	*Less than c.f.*	*Deviation from median* $\|d\|$	$f\|d\|$
0–10	5	18	18	19	342
10–20	15	16	34	9	144
20–30	25	15	49	1	15
30–40	35	12	61	11	132
40–50	45	10	71	21	210
50–60	55	5	76	31	155
60–70	65	2	78	41	82
70–80	75	2	80	51	102
		80			1182

$$\text{Median} = \text{the size of } \frac{80}{2} \text{ th item}$$

$$= 20 + \frac{6}{15} \times 10 = 24$$

$$\text{and then, mean deviation} = \frac{\Sigma f\,|d|}{\Sigma f}$$

$$= \frac{1182}{80} = 14.775.$$

Merits and demerits of the mean deviation

Merits

- It is easy to understand.
- As compared to standard deviation (discussed later), its computation is simple.
- As compared to standard deviation, it is less affected by extreme values.
- Since it is based on all values in the distribution, it is better than range or quartile deviation.

Demerits

- It lacks those algebraic properties which would facilitate its computation and establish its relation to other measures.
- Due to this, it is not suitable for further mathematical processing.

Coefficient of mean deviation

The coefficient or relative dispersion is found by dividing the mean deviations recorded. Thus,

$$\text{Coefficient of M.D.} = \frac{\text{Mean Deviation}}{\text{Mean}} \qquad ...(5)$$

(when deviations were recorded from the mean)

$$= \frac{\text{M.D.}}{\text{Median}} \qquad ...(6)$$

(when deviations were recorded from the median)

NOTES

Applying the above formula to Example 1.22.

$$\text{Coefficient of Mean deviation} = \frac{14.775}{24} = 0.616$$

(vii) Standard deviation

By far the most universally used and the most useful measure of dispersion is the standard deviation or root mean square deviation about the mean. We have seen that all the methods of measuring dispersion so far discussed are not universally adopted for want of adequacy and accuracy. The range is not satisfactory as its magnitude is determined by most extreme cases in the entire group. Further, the range is notable because it is dependent on the item whose size is largely matter of chance. Mean deviation method is also an unsatisfactory measure of scatter, as it ignores the algebraic signs of deviation. We desire a measure of scatter which is free from these shortcomings. To some extent standard deviation is one such measure.

The calculation of standard deviation differs in the following respects from that of mean deviation. First, in calculating standard deviation, the deviations are squared. This is done so as to get rid of negative signs without committing algebraic violence. Further, the squaring of deviations provides added weight to the extreme items, a desirable feature for certain types of series.

Secondly, the deviations are always recorded from the arithmetic mean, because although the sum of deviations is the minimum from the median, the sum of squares of deviations is minimum when deviations are measured from the arithmetic average. The deviation from $\bar{x}$ is represented by d.

Thus, standard deviation, σ (sigma) is defined as the square root of the mean of the squares of the deviations of individual items from their arithmetic mean.

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}} \qquad ...(7)$$

For grouped data (discrete variables)

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} \qquad ...(8)$$

and, for grouped data (continuous variables)

$$\sigma = \sqrt{\frac{\Sigma f(M-\bar{x})}{\Sigma f}} \qquad ...(9)$$

where M is the mid-value of the group.

The use of these formulae is illustrated by the following examples.

Example 1.23: Compute the standard deviation for the following data:

11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21.

Solution:

Here formula (7) is appropriate. We first calculate the mean as $\bar{x} = \sum x/N = 176/11 = 16$, and then calculate the deviation as follows:

x	$(x-\bar{x})$	$(x-\bar{x})^2$
11	–5	25
12	–4	16
13	–3	9
14	–2	4
15	–1	1
16	0	0
17	+1	1
18	+2	4
19	+3	9
20	+4	16
21	+5	25
176		10

Thus by formula (7).

$$\sigma = \sqrt{\frac{110}{11}} = \sqrt{10} = 3.16$$

Example 1.24: Find the standard deviation of the data in the following distributions:

x	12	13	14	15	16	17	18	20
f	4	11	32	21	15	8	6	4

Solution:

For this discrete variable grouped data, we use formula 8. Since for calculation of $\bar{x}$, we need $\sum fx$ and then for σ we need $\sum f(x-\bar{x})^2$, the calculations are conveniently made in the following format.

x	f	fx	$d = x - \bar{x}$	d^2	fd^2
12	4	48	–3	9	36
13	11	143	–2	4	44
14	32	448	–1	1	32
15	21	315	0	0	0
16	15	240	1	1	15
17	8	136	2	4	32
18	5	90	3	9	45
20	4	80	5	25	100
	100	1500			304

Here $\bar{x} = \sum fx/\sum f = 1500/100 = 15$

and $\sigma = \sqrt{\dfrac{\sum fd^2}{\sum f}}$

$= \sqrt{\dfrac{304}{100}} = \sqrt{3.04} = 1.74$

Example 1.25: Calculate the standard deviation of the following data.

Class	1–3	3–5	5–7	7–19	9–11	11–13	13–15
frequency	1	9	25	35	17	10	3

Solution: This is an example of continuous frequency series and formula 9 seems appropriate.

Class	*Mid-point* x	*Frequency* f	fx	*Deviation of mid-point x from mean (8)*	*Squared deviation* d^2	*Squared deviation times frequency* d^2
1–3	2	1	2	–6	36	36
3–5	4	9	36	–4	16	144
5–7	6	25	150	–2	4	100
7–9	8	35	280	0	0	0
9–11	10	17	170	2	4	68
11–13	12	10	120	4	16	160
13–15	14	3	42	6	36	108
		100	800			616

First the mean is calculated as

$$\bar{x} = \sum fx/\sum x = 800/100 = 8.0$$

Then the deviations are obtained from 8.0. The standard deviation

$$\sigma = \sqrt{\frac{\sum f(M-\bar{x})^2}{\sum f}}$$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f}} = \sqrt{\frac{616}{100}}$$

$$= 2.48$$

(a) Calculation of Standard Deviation by Short-cut Method

The three examples worked out above have one common simplifying feature, namely $\bar{x}$ in each, turned out to be an integer, thus, simplifying calculations. In most cases, it is very unlikely that it will turn out to be so. In such cases, the calculation of d and d^2 becomes quite time-consuming. Short-cut methods have consequently been developed. These are on the same lines as those for calculation of mean itself.

In the short-cut method, we calculate deviations x' from an assumed mean A.

NOTES

Then,

for ungrouped data

$$\sigma = \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2} \quad ...(10)$$

and for grouped data

$$\sigma = \sqrt{\frac{\sum fx'^2}{\sum f} - \left(\frac{fx'}{\sum f}\right)^2} \quad ...(11)$$

This formula is valid for both discrete and continuous variables. In case of continuous variables, x in the equation $x' = x - A$ stands for the mid-value of the class in question.

Note that the second term in each of the formulae is a correction term because of the difference in the values of A and $\bar{x}$. When A is taken as $\bar{x}$ itself, this correction is automatically reduced to zero. Examples 1.16 to 1.20 explain the use of these formulae.

Example 1.26: Compute the standard deviation by the short-cut method for the following data:

11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21

Solution: Let us assume that $A = 15$.

	$x' = (x - 15)$	x'^2
11	–4	16
12	–3	9
13	–2	4
14	–1	1
15	0	0
16	1	1
17	2	4
18	3	9
19	4	16
20	5	25
21	6	36
$N = 11$	$\sum x' = 11$	$\sum x'^2 = 121$

$$\sigma = \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2}$$

$$= \sqrt{\frac{121}{11} - \left(\frac{11}{11}\right)^2}$$

$$= \sqrt{11 - 1}$$

$$= \sqrt{10}$$

$$= 3.16.$$

Another method

If we assumed A as zero, then the deviation of each item from the assumed mean is the same as the value of item itself. Thus, 11 deviates from the assumed mean of zero by 11, 12 deviates by 12, and so on. As such, we work with deviations without having to compute them, and the formula takes the following shape:

x	x^2
11	121
12	144
13	169
14	196
15	225
16	256
17	289
18	324
19	361
20	400
21	441
176	2,926

$$\sigma = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}$$

$$= \sqrt{\frac{2926}{11} - \left(\frac{176}{11}\right)^2} = \sqrt{266 - 256} = 3.16$$

Example 1.27: Calculate the standard deviation of the following data by short method.

Person	1	2	3	4	5	6	7
Monthly income (Rupees)	300	400	420	440	460	480	580

Solution: In this data, the values of the variable are very large making calculations cumbersome. It is advantageous to take a common factor out. Thus, we use $x' = \frac{x - A}{20}$. The standard deviation is calculated using x' and then the true value of σ is obtained by multiplying back by 20. The effective formula then is

$$\sigma = C \times \sqrt{\frac{\Sigma x'^2}{N} - \left(\frac{\Sigma x'}{N}\right)^2}$$

where C represents the common factor.

Using $x' = (x - 420)/20$.

NOTES

x	Deviation from Assumed mean $x' = (x - 420)$	x'	x'^2
300	–120	–6	36
400	–20	–1	1
420	0	0	0
		–7	
440	20	1	1
460	40	2	4
480	60	3	9
580	160	8	64
		+ 14	
$N = 7$		7	115

$$\sigma = 20 \times \sqrt{\frac{\Sigma x'^2}{N} - \left(\frac{\Sigma x'}{N}\right)^2}$$

$$= 20\sqrt{\frac{115}{7} - \left(\frac{7}{7}\right)^2}$$

$$= 78.56$$

Example 1.28: Calculate the standard deviation from the following data:

Size	6	9	12	15	18
Frequency	7	12	19	10	2

Solution:

x	Frequency f	Deviation from assumed mean 12	Deviation divided by common factor 3 x'	x' times frequency fx'	x'^2 times frequency fx'^2
6	7	–6	–2	–14	28
9	12	–3	–1	–12	12
12	19	0	0	0	0
15	10	3	1	10	10
18	2	6	2	4	8
	$N = 50$			$\Sigma fx' = -12$	$\Sigma fx'^2 = 58$

Since deviations have been divided by a common factor, we use

$$\sigma = C\sqrt{\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2}$$

$$= 3\sqrt{\frac{58}{50} - \left(\frac{-12}{50}\right)^2}$$

$$= 3\sqrt{1.1600 - .0576} = 3 \times 1.05 = 3.15.$$

Example 1.29: Obtain the mean and standard deviation of the first N natural numbers, i.e., of 1, 2, 3, ..., N – 1, N.

Solution: Let x denote the variable which assumes the values of the first N natural numbers.

Then

$$\bar{x} = \frac{\sum_1^N x}{N} = \frac{\frac{N(N+1)}{2}}{N} = \frac{N+1}{2}$$

because $\sum_1^N x = 1 + 2 + 3 + ... + (N-1) + N$

$$= \frac{N(N+1)}{2}$$

To calculate the standard deviation σ, we use 0 as the assumed mean A. Then

$$\sigma = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}$$

But $\Sigma x^2 = 1^2 + 2^2 + 3^2 + ... (N-1)^2 + N^2 = \frac{N(N+1)(2N+1)}{6}$

Therefore

$$\sigma = \sqrt{\frac{N(N+1)(2N+1)}{6N} - \frac{N^2(N+1)^2}{4N^2}}$$

$$= \sqrt{\frac{(N+1)}{2}\left[\frac{2N+1}{3} - \frac{N+1}{2}\right]} = \sqrt{\frac{(N+1)(N-1)}{12}}$$

Thus for first 11 natural numbers

$$\bar{x} = \frac{11+1}{2} = 6$$

and $\sigma = \sqrt{\frac{(11+1)(11-1)}{12}} = \sqrt{10} = 3.16$

Example 1.30:

	Mid-point x	*Frequency* f	*Deviation from class of assumed mean* x'	*Deviation time frequency* fx'	*Squared deviation times frequency* fx'^2
0–10	5	18	–2	–36	72
10–20	15	16	–1	–16	16
				–52	
20–30	25	15	0	0	0
30–40	35	12	1	12	12
40–50	45	10	2	20	40
50–60	55	5	3	15	45
60–70	65	2	4	8	32
70–80	75	1	5	5	25
				– 60	
		79		60	242
				–52	
				$\Sigma fx' = 8$	

NOTES

Solution: Since the deviations are from assumed mean and expressed in terms of class-interval units,

$$\sigma = i \times \sqrt{\frac{\Sigma x'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2}$$

$$= 10 \times \sqrt{\frac{242}{79} - \left(\frac{8}{79}\right)^2}$$

$$= 10 \times 1.75 = 17.5.$$

(b) Combining Standard Deviations of Two Distributions

If we were given two sets of data of N_1 and N_2 items with means $\bar{x}_1$ and $\bar{x}_2$ and standard deviations σ_1 and σ_2 respectively, we can obtain the mean and standard deviation $\bar{x}$ and σ of the combined distribution by the following formulae:

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} \quad ...(12)$$

and $$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}} \quad ...(13)$$

Example 1.31: The mean and standard deviations of two distributions of 100 and 150 items are 50, 5 and 40, 6 respectively. Find the standard deviation of all taken together.

Solution: Combined mean

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} = \frac{100 \times 50 + 150 \times 40}{100 + 150}$$

$$= 44$$

Combined standard deviation

$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}}$$

$$= \sqrt{\frac{100 \times (5)^2 + 150\,(6)^2 + 100\,(44 - 50)^2 + 150\,(44 - 40)^2}{100 + 150}}$$

$$= 7.46.$$

Example 1.32: A distribution consists of three components with 200, 250, 300 items having mean 25, 10 and 15 and standard deviation 3, 4 and 5, respectively. Find the standard deviation of the combined distribution.

Solution: In the usual notations, we are given here

$N_1 = 200,\ N_2 = 250,\ N_3 = 300$

$\bar{x}_1 = 25,\ \bar{x}_2 = 10,\ \bar{x}_3 = 15$

The formulae (12) and (13) can easily be extended for combination of three series as

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + N_3\bar{x}_3}{N_1 + N_2 + N_3}$$

$$= \frac{200 \times 25 + 250 \times 10 + 300 \times 15}{200 + 250 + 300}$$

$$= \frac{12000}{750} = 16$$

and

NOTES

$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2 + N_3(\bar{x} - \bar{x}_3)^2}{N_1 + N_2 + N_3}}$$

$$= \sqrt{\frac{200 \times 9 + 250 \times 16 + 300 \times 25 + 200 \times 81 + 250 \times 36 + 300 \times 1}{200 + 250 + 300}}$$

$$= \sqrt{51.73} = 7.19.$$

(c) Comparison of Various Measures of Dispersion

The range is the easiest to calculate the measure of dispersion, but since it depends on extreme values, it is extremely sensitive to the size of the sample, and to the sample variability. In fact, as the sample size increases the range increases dramatically, because the more the items one considers, the more likely it is that some item will turn up which is larger than the previous maximum or smaller than the previous minimum. So, it is, in general, impossible to interpret properly the significance of a given range unless the sample size is constant. It is for this reason that there appears to be only one valid application of the range, namely in statistical quality control where the same sample size is repeatedly used, so that comparison of ranges are not distorted by differences in sample size.

The quartile deviations and other such positional measures of dispersions are also easy to calculate but suffer from the disadvantage that they are not amenable to algebraic treatment. Similarly, the mean deviation is not suitable because we cannot obtain the mean deviation of a combined series from the deviations of component series. However, it is easy to interpret and easier to calculate than the standard deviation.

The standard deviation of a set of data, on the other hand, is one of the most important statistics describing it. It lends itself to rigorous algebraic treatment, is rigidly defined and is based on all observations. It is, therefore, quite insensitive to sample size (provided the size is 'large enough') and is least affected by sampling variations.

It is used extensively in testing of hypothesis about population parameters based on sampling statistics.

In fact, the standard deviations has such stable mathematical properties that it is used as a standard scale for measuring deviations from the mean. If we are told that the performance of an individual is 10 points better than the mean, it really does not tell us enough, for 10 points may or may not be a large enough difference to be of significance. But if we know that the s for the score is only 4 points, so that on this scale, the performance is 2.5s better than the mean, the statement becomes meaningful. This indicates an extremely good performance.

NOTES

This sigma scale is a very commonly used scale for measuring and specifying deviations which immediately suggest the significance of the deviation.

The only disadvantages of the standard deviation lies in the amount of work involved in its calculation, and the large weight it attaches to extreme values because of the process of squaring involved in its calculations.

(viii) Coefficient of Variation

The square of standard deviation, namely σ^2, is termed as variance and is more often specified than the standard deviation. Clearly, it has the same properties as standard deviation.

As is clear, the standard deviation σ or its square, the variance, cannot be very useful in comparing two series where either the units are different or the mean values are different. Thus, a σ of 5 on an examination where the mean score is 30 has an altogether different meaning than on an examination where the mean score is 90. Clearly, the variability in the second examination is much less. To take care of this problem, we define and use a coefficient of variation, V,

$$V = \frac{\sigma}{\bar{x}} \times 100$$

expressed as percentage.

Example 1.33: The following are the scores of two batsmen A and B in a series of innings:

A	12	115	6	73	7	19	119	36	84	29
B	47	12	76	42	4	51	37	48	13	0

Who is the better run-getter? Who is more consistent?

Solution: In order to decide as to which of the two batsmen, A and B, is the better run-getter, we should find their batting averages. The one whose average is higher will be considered as a better batsman.

To determine the consistency in batting we should determine the coefficient of variation. The less this coefficient the more consistent will be the player.

A			*B*		
Score x	x	x^2	*Scores* x	x	x^2
12	–38	1,444	47	14	196
115	+65	4,225	12	–21	441
6	–44	1,936	76	43	1,849
73	+23	529	42	9	81
7	–43	1,849	–4	– 29	841
19	–31	961	51	18	324
119	+69	4,761	37	4	16
36	–14	196	48	15	225
84	+34	1,156	13	–20	400
29	–21	441	0	–33	1,089
$\sum x = 500$		17,498	$\sum x = 330$		5,462

Batsman *A*:

$$\bar{x} = \frac{500}{10} = 50$$

$$\sigma = \sqrt{\frac{17,498}{10}} = 41.83$$

$$V = \frac{41.83 \times 100}{50}$$

$$= 83.66 \text{ per cent}$$

Batsman *B*:

$$\bar{x} = \frac{330}{10} = 33$$

$$\sigma = \sqrt{\frac{5,462}{10}} = 23.37$$

$$V = \frac{23.37}{33} \times 100$$

$$= 70.8 \text{ per cent}$$

A is a better batsman since his average is 50 as compared to 33 of *B*. But *B* is more consistent since the variation in his case is 70.8 as compared to 83.66 of *A*.

Example 1.34: The following table gives the age distribution of students admitted to a college in the years 1914 and 1918. Find which of the two groups is more variable in age.

Age	*Number of students in*	
	1914	*1918*
15	–	1
16	1	6
17	3	34
18	8	22
19	12	35
20	14	20
21	13	7
22	5	19
23	2	3
24	3	–
25	1	–
26	–	–
27	1	–

Solution:

Age	*Assumed Mean–21* *1914*				*Assumed Mean–19* *1918*			
	f	*x'*	*fx'*	*fx'²*	*f*	*x'*	*fx*	*fx'²*
15	0	–6	0	0	1	–4	–4	16
16	1	–5	–5	25	6	–3	–18	54
17	3	–4	–12	48	34	–2	–68	136
18	8	–3	–24	72	22	–1	–22	22
19	12	–2	–24	48			–112	
20	14	–1	–14	14				
			–79		35	0	0	0
21	13	0	0	0	20	1	20	20
22	5	1	5	5	7	2	14	28
23	2	2	4	8	19	3	57	171
24	3	3	9	27	3	4	12	48
25	1	4	4	16	147		+103	495
26	0	5	0	0			–9	
27	1	6	6	36				
	63		+28	299				
			–51					

NOTES

1914 Group:

$$\sigma = \sqrt{\frac{\sum fx'^2}{N} - \left[\frac{\sum(fx')}{N}\right]^2}$$

$$= \sqrt{\frac{299}{63} - \left(\frac{-51}{63}\right)^2}$$

$$= \sqrt{4.476 - 0.655} = \sqrt{4.091}$$

$$= 2.02.$$

$$\bar{x} = 21 + \left(\frac{-51}{63}\right) = 21 - 8 = 20.2$$

$$V = \frac{2.02}{20.2} \times 100$$

$$= \frac{202}{20.2} = 10$$

1918 Group:

$$\sigma = \sqrt{\frac{495}{147} - \left(\frac{-9}{147}\right)^2} = \sqrt{3.3673 - 0.0037}$$

$$= \sqrt{3.3636} = 1.834$$

$$\bar{x} = 19 + \left(\frac{-9}{147}\right)$$

$$= 19 - .06 = 18.94$$

$$V = \frac{1.834}{18.94} \times 100$$

$$= 9.68$$

The coefficient of variation of the 1914 group is 10 and that of the 1918 group 9.68. This means that the 1914 group is more variable, but only barely so.

Example 1.35: You are supplied the following data about the height of boys and girls studying in a college.

	Boys	Girls
Number	72	38
Average height (inches)	68	61
Variance of distribution	9	4

You are required to find out:

(a) In which sex, boys or girls, is there greater variability in individual heights.

(b) Common average height in boys and girls.

(c) Standard deviation of the height of boys and girls taken together.

(d) Combined variability.

Solution:

(*a*) C.V. of boys' height $= \frac{\sigma_1}{\bar{x}_1} \times 100 = \frac{\sqrt{9}}{68} \times 100 = 4.41\%$

C.V. of girls' height $= \frac{\sigma_2}{\bar{x}_2} \times 100 = \frac{\sqrt{4}}{61} \times 100 = 3.28\%$

Thus there is a greater variability in the height of boys than that of the girls.

(*b*) Height of boys and girls combined is

$$\bar{x}_{12} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2}$$

$$= \frac{72 \times 68 + 38 \times 61}{72 + 38} = \frac{7214}{110} = 65.58 \text{ inches approx.}$$

(*c*) The combined standard deviation may be calculated by applying the following formula:

$$\sigma_{12}^2 = \frac{N_1\sigma_1^2 + N_2\sigma_2^2}{N_1 + N_2} + \frac{N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}$$

$$= \frac{72 \times 9 + 38 \times 4}{72 + 38} + \frac{72\,(65.58 - 68)^2 + 38\,(65.58 - 61)^2}{72 + 38}$$

$$= \frac{2018.794}{110} = 18.35$$

$$\sigma_{12} = 4.28 \text{ inches.}$$

(*d*) Combined variability $= \frac{\sigma}{\bar{x}} \times 100 = \frac{4.28}{65.58} \times 100 = 6.53$

1.5.2 Simple Correlation and Regression

Correlation analysis is the statistical tool generally used to describe the degree to which, one variable is related to another. The relationship, if any, is usually assumed to be a linear one. This analysis is used quite frequently in conjunction with regression analysis to measure how well the regression line explains the variations of the dependent variable. In fact, the word correlation refers to the relationship or interdependence between two variables. There are various phenomenons which have relation to each other. For instance, when demand of a certain commodity increases, then its price goes up and when its demand decreases, its price comes down. Similarly, with age the height of the children, with height the weight of the children, with money the supply and the general level of prices go up. Such sort of relationship can as well be noticed for several other phenomena. The theory by means of which quantitative connections between two sets of phenomena are determined is called the '*Theory of Correlation*'.

On the basis of the theory of correlation, one can study the comparative changes occurring in two related phenomena and their cause-effect relation can be examined. It should, however, be borne in mind that relationship like 'black cat causes bad luck', 'filled up pitchers result in good fortune' and similar other beliefs of the people cannot be explained by the theory of correlation, since they are all imaginary and are incapable of being justified mathematically. Thus, correlation is concerned with relationship between two related and quantifiable variables. If two quantities vary in sympathy, so that a movement (an increase or decrease) in one, tends to be accompanied by a movement in the same or opposite direction in the other and the greater the change in the one, the greater is the change in the other,

NOTES

the quantities are said to be correlated. This type of relationship is known as correlation or what is sometimes called, in statistics, as covariation.

For correlation, it is essential that the two phenomena should have cause-effect relationship. If such relationship does not exist then one should not talk of correlation. For example, if the height of the students as well as the height of the trees increases, then one should not call it a case of correlation because the two phenomena, viz., the height of students and the height of trees are not even casually related. But, the relationship between the price of a commodity and its demand, the price of a commodity and its supply, the rate of interest and savings, etc. are examples of correlation, since in all such cases the change in one phenomenon is explained by a change in other phenomenon.

It is appropriate here to mention that correlation in case of phenomena pertaining to natural sciences can be reduced to absolute mathematical term, e.g., heat always increases with light. But in phenomena pertaining to social sciences it is often difficult to establish any absolute relationship between two phenomena. Hence, in social sciences, we must take the fact of correlation being established if in a large number of cases, two variables always tend to move in the same or opposite direction.

Correlation can either be positive or it can be negative. Whether correlation is positive or negative would depend upon the direction in which the variables are moving. If both variables are changing in the same direction, then correlation is said to be positive, but, when the variations in the two variables take place in opposite direction, the correlation is termed as negative.

Table 1.9 *Nature of Correlation*

Changes in Independent Variable	*Changes in Dependent Variable*	*Nature of Correlation*
Increase (+)↑	Increase (+)↑	Positive (+)
Decrease (–)↓	Decrease (–)↓	Positive (+)
Increase (+)↑	Decrease (–)↓	Negative (–)
Decrease (–)↓	Increase (+)↑	Negative (–)

Statisticians have developed *two measures for describing the correlation* between two variables, viz., the coefficient of determination and the coefficient of correlation. We now explain, illustrate and interpret the said two coefficients concerning the relationship between two variables as under:

(i) Coefficient of Determination

The coefficient of determination (symbolically indicated as r^2, though some people would prefer to put it as R^2) is a measure of the degree of linear association or correlation between two variables, say X and Y, one of which happens to be independent variable and the other being dependent variable. This coefficient is based on the following two kinds of variations:

(*i*) The variation of the Y values around the fitted regression line viz., $\Sigma\left(Y-\hat{Y}\right)^2$, technically known as the unexplained variation.

NOTES

(*ii*) The variation of the Y values around their own mean viz., $\Sigma(Y-\overline{Y})^2$, technically known as the total variation.

If we subtract the unexplained variation from the total variation, we obtain what is known as the explained variation, i.e., the variation explained by the line of regression. Thus, explained Variation = (Total variation) – (Unexplained variation)

$$= \Sigma(Y-\overline{Y})^2 - \Sigma(Y-\hat{Y})^2$$

$$= \Sigma(\hat{Y}-\overline{Y})^2$$

The Total and Explained as well as Unexplained variations can be shown as given in Figure 1.9.

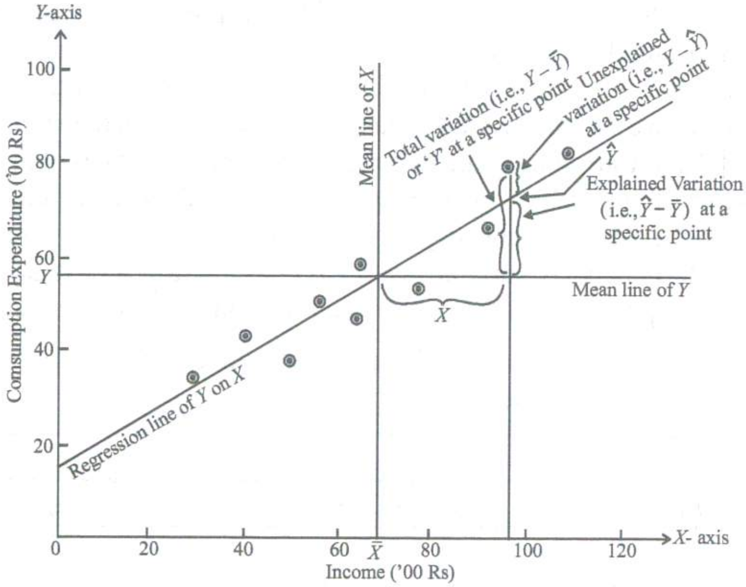


Fig. 1.9 *Diagram Showing Total, Explained and Unexplained Variations*

Coefficients of determination is that fraction of the total variation of Y which is explained by the regression line. In other words, coefficient of determination is the ratio of explained variation to total variation in the Y variable related to the X variable. Coefficient of determination algebraically can be stated as under:

$$r^2 = \frac{\text{Explained variation}}{\text{Total variation}}$$

$$= \frac{\Sigma(\hat{Y}-\overline{Y})^2}{\Sigma(Y-\overline{Y})^2}$$

Alternatively r^2 can also be stated as under:

$$r^2 = 1 - \frac{\text{Explained variation}}{\text{Total variation}}$$

$$= 1 - \frac{\Sigma(\hat{Y}-\overline{Y})^2}{\Sigma(Y-\overline{Y})^2}$$

Interpreting r^2

The coefficient of determination can have a value ranging from zero to one. The value of one can occur only if the unexplained variation is zero, which simply means that all the data points in the Scatter diagram fall exactly on the regression line. For a zero value to occur, $\Sigma(Y-\bar{Y})^2 = \Sigma(Y-\hat{Y})^2$, which simply means that X tells us nothing about Y and hence there is no regression relationship between X and Y variables. Values between 0 and 1 indicate the 'Goodness of fit' of the regression line to the sample data. The higher the value of r^2, the better the fit. In other words, the value of r^2 will lie somewhere between 0 and 1. If r^2 has a zero value then it indicates no correlation but if it has a value equal to 1 then it indicates that there is perfect correlation and as such the regression line is a perfect estimator. But in most of the cases, the value of r^2 will lie somewhere between these two extremes of 1 and 0. One should remember that r^2 close to 1 indicates a strong correlation between X and Y while an r^2 near zero means there is little correlation between these two variables. r^2 value can as well be interpreted by looking at the amount of the variation in Y, the dependant variable, that is explained by the regression line. Supposing, we get a value of $r^2 = 0.925$ then this would mean that the variations in independent variable (say X) would explain 92.5 per cent of the variation in the dependent variable (say Y). If r^2 is close to 1 then it indicates that the regression equation explains most of the variations in the dependent variable.

Example 1.36: Calculate the coefficient of determination (r^2) using data given below. Calculate and analyse the result.

Observations	1	2	3	4	5	6	7	8	9	10
Income (X) ('00 ₹)	41	65	50	57	96	94	110	30	79	65
Consumption Expenditure (Y) ('00 ₹)	44	60	39	51	80	68	84	34	55	48

Solution: r^2 can be worked out as shown below:

Since, $$r^2 = 1 - \frac{\text{Unexplained variation}}{\text{Total variation}} = 1 - \frac{\Sigma\left(Y-\hat{Y}\right)^2}{\Sigma\left(Y-\bar{Y}\right)^2}$$

As, $\Sigma\left(Y-\bar{Y}\right)^2 = \Sigma Y^2 = \left(\Sigma Y^2 - n\bar{Y}^2\right)$, we can write,

$$r^2 = 1 - \frac{\Sigma\left(Y-\hat{Y}\right)^2}{\Sigma Y^2 - n\bar{Y}^2}$$

Calculating and putting the various values, we have the following equation:

$$r^2 = 1 - \frac{260.54}{34223 - 10(56.3)^2} = 1 - \frac{260.54}{2526.10} = 0.897$$

Analysis of the result: The regression equation used to calculate the value of the coefficient of determination (r^2) from the sample data shows that, about 90 per cent of the variations in consumption expenditure can be explained. In other words, it means that the variations in income explain about 90 per cent of variations in consumption expenditure.

NOTES

Observations	1	2	3	4	5	6	7	8	9	10
Income (X) ('00 ₹)	41	65	50	57	96	94	110	30	79	65
Consumption Expenditure (Y) ('00 ₹)	44	60	39	51	80	68	84	34	55	48

(ii) Coefficient of Correlation

The coefficient of correlation, symbolically denoted by 'r', is another important measure to describe how well one variable is explained by another. It measures the degree of relationship between the two casually related variables. The value of this coefficient can never be more than +1 or less than –1. Thus, +1 and –1 are the limits of this coefficient. For a unit change in independent variable, if there happens to be a constant change in the dependent variable in the same direction, then the value of the coefficient will be +1 indicative of the perfect positive correlation; but if such a change occurs in the opposite direction, the value of the coefficient will be –1, indicating the perfect negative correlation. In practical life, the possibility of obtaining either a perfect positive or perfect negative correlation is very remote particularly in respect of phenomena concerning social sciences. If the coefficient of correlation has a zero value then it means that there exists no correlation between the variables under study.

There are several methods of finding the coefficient of correlation but the following ones are considered important:

(*a*) Coefficient of correlation by the Method of Least Squares.

(*b*) Coefficient of correlation using Simple Regression Coefficients.

(*c*) Coefficient of correlation through Product Moment Method or Karl Pearson's Coefficient of correlation.

Whichever of these above mentioned three methods we adopt, we get the same value of r.

(*a*) Coefficient of correlation by the Method of Least Squares

Under this method, first of all, the estimating equation is obtained using least square method of simple regression analysis. The equation is worked out as,

$$\hat{Y} = a + bX_i$$

Total variation $= \Sigma\left(Y - \overline{Y}\right)^2$

Unexplained variation $= \Sigma\left(Y - \hat{Y}\right)^2$

Explained variation $= \Sigma\left(\hat{Y} - \overline{Y}\right)^2$

Then, by applying the following formulae, we can find the value of the coefficient of correlation:

$$r = \sqrt{r^2} = \sqrt{\frac{\text{Explained variation}}{\text{Total variation}}}$$

$$= \sqrt{1 - \frac{\text{Unexplained variation}}{\text{Total variation}}}$$

$$= \sqrt{1 - \frac{\Sigma\left(Y - \hat{Y}\right)^2}{\Sigma\left(Y - \overline{Y}\right)^2}}$$

This clearly shows that *the coefficient of correlation happens to be the squareroot of the coefficient of determination.*

Short-cut formula for finding the value of 'r' by the method of least squares can be repeated and readily written as follows:

$$r = \sqrt{\frac{a\Sigma Y + b\Sigma XY - n\overline{Y}^2}{\Sigma Y^2 - n\overline{Y}^2}}$$

Where,
a = Y-intercept
b = Slope of the estimating equation
X = Values of the independent variable
Y = Values of dependent variable
$\overline{Y}$ = Mean of the observed values of Y
n = Number of items in the sample (i.e., pairs of observed data)

The plus (+) or the minus (–) sign of the coefficient of correlation worked out by the method of least squares, is related to the sign of 'b' in the estimating equation viz., $\hat{Y} = a + bX_i$. If 'b' has a minus sign, the sign of 'r' will also be minus but if 'b' has a plus sign, then the sign of 'r' will also be plus. The value of 'r' indicates the degree along with the direction of the relationship between the two variables X and Y.

(*b*) Coefficient of correlation using Simple Regression Coefficients

Under this method, the estimating equation of Y and the estimating equation of X is worked out using the method of least squares. From these estimating equations we find the regression coefficient of X on Y, i.e., the slope of the estimating equation of X (symbolically written as b_{XY}) and this happens to be equal to $r\frac{\sigma_X}{\sigma_Y}$ and similarly, we find the regression coefficient of Y on X, i.e., the slope of the estimating equation of Y (symbolically written as b_{YX}) and this happens to be equal to $r\frac{\sigma_Y}{\sigma_X}$. For finding '$r$', the square root of the product of these two regression coefficients are work out as following:

$$r = \sqrt{b_{XY}.b_{YX}}$$

$$= \sqrt{r\frac{\sigma_X}{\sigma_Y}.r\frac{\sigma_Y}{\sigma_X}}$$

$$= \sqrt{r^2} = r$$

As stated earlier, the sign of 'r' will depend upon the sign of the regression coefficients. If they have minus sign, then 'r' will take minus sign but the sign of 'r' will be plus if regression coefficients have plus sign.

NOTES

(c) Karl Pearson's Coefficient

Karl Pearson's method is most widely used method of measuring the relationship between two variables. This coefficient is based on the following assumptions:

(*i*) There is a linear relationship between the two variables which means that straight line would be obtained if the observed data are plotted on a graph.

(*ii*) The two variables are casually related which means that one of the variables is independent and the other one is dependent.

(*iii*) A large number of independent causes are operating in both the variables so as to produce a normal distribution.

According to Karl Pearson, 'r' can be worked out as under:

$$r = \frac{\Sigma XY}{n\sigma_X \sigma_Y}$$

Where,

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

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

σ_X = Standard deviation of X series and is equal to $\sqrt{\frac{\Sigma X^2}{n}}$

σ_Y = Standard deviation of Y series and is equal to $\sqrt{\frac{\Sigma Y^2}{n}}$

n = Number of pairs of X and Y observed.

A short-cut formula, known as the Product Moment Formula, can be derived from the above stated formula as under:

$$r = \frac{\Sigma XY}{n\sigma_X \sigma_Y}$$

$$= \frac{\Sigma XY}{\sqrt{\frac{\Sigma X^2}{n} \cdot \frac{\Sigma Y^2}{n}}}$$

$$n = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

The above formulae are based on obtaining true means (viz. $\bar{X}$ and $\bar{Y}$) first and then doing all other calculations. This happens to be a tedious task, particularly if the true means are in fractions. To avoid difficult calculations, we make use

NOTES

of the assumed means in taking out deviations and doing the related calculations. In such a situation, we can use the following formula for finding the value of 'r':[2]

(*i*) *In case of ungrouped data:*

$$r = \frac{\frac{\Sigma dX.dY}{n} - \left(\frac{\Sigma dX}{n}.\frac{\Sigma dY}{n}\right)}{\sqrt{\frac{\Sigma dX^2}{n} - \left(\frac{\Sigma dX}{n}\right)^2}.\sqrt{\frac{\Sigma dY^2}{n} - .\left(\frac{\Sigma dY}{n}\right)^2}}$$

$$= \frac{\Sigma dX.dY - \left(\frac{\Sigma dX \times \Sigma dY}{n}\right)}{\sqrt{\Sigma dX^2 - \frac{(\Sigma dX)^2}{n}}\sqrt{\Sigma dY^2 - \frac{(\Sigma dY)^2}{n}}}$$

Where, $\Sigma dX = \Sigma(X - X_A)$ X_A = Assumed average of X

$\Sigma dY = \Sigma(Y - Y_A)$ Y_A = Assumed average of Y

$\Sigma dX^2 = \Sigma(X - X_A)^2$

$\Sigma dY^2 = \Sigma(Y - Y_A)^2$

$\Sigma dX \,.\, dY = \Sigma(X - X_A)\,(Y - Y_A)$

n = Number of pairs of observations of X and Y

(*ii*) *In case of grouped data:*

$$r = \frac{\frac{\Sigma\, fdX.dY}{n} - \left(\frac{\Sigma\, fdX}{n}.\frac{\Sigma\, fdY}{n}\right)}{\sqrt{\frac{\Sigma\, fdX^2}{n} - \left(\frac{\Sigma\, fdX}{n}\right)^2}\sqrt{\frac{\Sigma\, fdY^2}{n} - \left(\frac{\Sigma\, fdY}{n}\right)^2}}$$

or $$r = \frac{\Sigma\, fdX.dY - \left(\frac{\Sigma\, fdX.\Sigma\, fdY}{n}\right)}{\sqrt{\Sigma\, fdX^2 - \left(\frac{\Sigma\, fdX}{n}\right)^2}\sqrt{\Sigma\, fdY^2 - \left(\frac{\Sigma\, fdY}{n}\right)^2}}$$

Where, $\Sigma fdX.dY = 0\Sigma f\,(X - X_A)\,(Y - Y_A)$

$\Sigma fdX = \Sigma f\,(X - X_A)$

$\Sigma fdY = \Sigma f\,(Y - Y_A)$

$\Sigma fdY^2 = \Sigma f\,(Y - Y_A)^2$

$\Sigma fdX^2 = \Sigma f\,(X - X_A)^2$

n = Number of pairs of observations of X and Y.

NOTES

Probable Error (P.E.) of the Coefficient of Correlation

Probable Error (P.E.) of r is very useful in interpreting the value of r and is worked out as under for Karl Pearson's coefficient of correlation:

$$\text{P.E.} = 0.6745\frac{1-r^2}{\sqrt{n}}$$

If r is less than its P.E., it is not at all significant. If r is more than P.E., there is correlation. *If r is more than 6 times its P.E. and greater than ± 0.5, then it is considered significant.*

Example 1.37:

From the following data calculate 'r' between X and Y applying the following three methods:

(*i*) The method of least squares.

(*ii*) The method based on regression coefficients.

(*iii*) The product moment method of Karl Pearson.

Verify the obtained result of any one method with that of another.

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

Solution:

Let us develop the following table for calculating the value of 'r':

X	Y	X^2	Y^2	XY
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
8	16	64	256	128
9	15	81	225	135
n=9				
$\Sigma X = 45$	$\Sigma Y = 108$	$\Sigma X^2 = 285$	$\Sigma Y^2 = 1356$	$\Sigma XY = 597$

$\therefore \quad \bar{X} = 5; \qquad \bar{Y} = 12$

(*i*) *Coefficient of correlation by the method of least squares is worked out as under:*

First of all find out the estimating equation,

$$\hat{Y} = a + bX_i$$

Where, $$b = \frac{\Sigma XY - n\overline{XY}}{\Sigma X^2 - n\overline{X}^2}$$

$$= \frac{597-9(5)(12)}{285-9(25)} = \frac{597-540}{285-225} = \frac{57}{60} = 0.95$$

NOTES

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

$$= 12 - 0.95(5) = 12 - 4.75 = 7.25$$

Hence, $$\hat{Y} = 7.25 + 0.95X_i$$

Now 'r' can be worked out as under by the method of least squares,

$$r = \sqrt{1 - \frac{\text{Unexplained variation}}{\text{Total variation}}}$$

$$= \sqrt{1 - \frac{\Sigma(Y-\hat{Y})^2}{\Sigma(Y-\overline{Y})^2}} = \sqrt{\frac{\Sigma(\hat{Y}-\overline{Y})^2}{\Sigma(Y-\overline{Y})^2}}$$

$$= \sqrt{\frac{a\Sigma Y + b\Sigma XY - n\overline{Y}^2}{\Sigma Y^2 - n\overline{Y}^2}}$$

This is as per short-cut formula,

$$r = \sqrt{\frac{7.25(108) + 0.95(597) - 9(12)^2}{1356 - 9(12)^2}}$$

$$= \sqrt{\frac{783 + 567.15 - 1296}{1356 - 1296}}$$

$$= \sqrt{\frac{54.15}{60}} = \sqrt{0.9025} = 0.95$$

(*ii*) *Coefficient of correlation by the method based on regression coefficients is worked out as under:*

$\because$ Regression coefficients of Y on X,

i.e., $$b_{YX} = \frac{\Sigma XY - n\overline{XY}}{\Sigma X^2 - n\overline{X}^2}$$

$$= \frac{597 - 9\times5\times12}{285 - 9(5)^2} = \frac{597-540}{285-225} = \frac{57}{60}$$

Regression coefficient of X on Y,

i.e., $$b_{XY} = \frac{\Sigma XY - n\overline{XY}}{\Sigma Y^2 - n\overline{Y}^2}$$

$$= \frac{597 - 9\times5\times12}{1356 - 9(12)^2} = \frac{597-540}{1356-1296} = \frac{57}{60}$$

Hence, $$r = \sqrt{b_{YX} \cdot b_{XY}}$$

$$= \sqrt{\frac{57}{60} \times \frac{57}{60}} = \frac{57}{60} = 0.95$$

(iii) *Coefficient of correlation by the product moment method of Karl Pearson is worked out as under:*

$$r = \frac{\Sigma XY - n\overline{X}\overline{Y}}{\sqrt{\Sigma X^2 - n\overline{X}^2}\sqrt{\Sigma Y^2 - n\overline{Y}^2}}$$

$$= \frac{597 - 9(5)(12)}{\sqrt{285 - 9(5)^2}\sqrt{1356 - 9(12)^2}}$$

$$= \frac{597 - 540}{\sqrt{285 - 225}\sqrt{1356 - 1296}} = \frac{57}{\sqrt{60}\sqrt{60}} = \frac{57}{60} = 0.95$$

Hence, we get the value of $r = 0.95$. We get the same value applying the other two methods also. Therefore, whichever method we apply, the results will be the same.

Some Other Measures

Two other measures are often talked about along with the coefficients of determinations and that of correlation. These are as follows:

(i) *Coefficient of Nondetermination.* Instead of using coefficient of determination, sometimes coefficient of nondetermination is used. Coefficient of nondetermination (denoted by k^2) is the ratio of unexplained variation to total variation in the Y variable related to the X variable. Algebrically, we can write it as follows:

$$k^2 = \frac{\text{Unexplained variation}}{\text{Total variation}} = \frac{\Sigma(Y - \hat{Y})^2}{\Sigma(Y - \overline{Y})^2}$$

Concerning the data of Example 1.36 of this unit, coefficient of nondetermination will be calculated as follows:

$$k^2 = \frac{260.54}{2526.10} = 0.103$$

The value of k^2 shows that about 10 per cent of the variation in consumption expenditure remains unexplained by the regression equation we had worked out, viz., $\hat{Y} = 14.000 + 0.616X_i$. In simple terms, this means that variable other than X is responsible for 10 per cent of the variations in the dependent variable Y in the given case.

Coefficient of nondetermination can as well be worked out as under:

$$k^2 = 1 - r^2$$

Accordingly for Example 1.36, it will be equal to 1–0.897 = 0.103

Note: *Always remember that* $r^2 + k^2 = 1$.

NOTES

(ii) *Coefficient of Alienation.* Based on k^2, we can work out one more measure namely the Coefficient of alienation, symbolically written as 'k'.

Thus, Coefficient of alienation, i.e., 'k' $=\sqrt{k^2}$

Unlike $r + k^2 = 1$, the sum of 'r' and 'k' will not be equal to 1 unless one of the two coefficients is 1 and in this case the remaining coefficients must be zero. In all other cases, 'r' + 'k' > 1. Coefficient of alienation is not a popular measure from practical point of view and is used very rarely.

(iii) Rank Correlation

If observations on two variables are given in the form of ranks and not as numerical values, it is possible to compute what is known as rank correlation between the two series.

Computation

The rank correlation, written ρ, is a descriptive index of agreement between ranks over individuals. It is the same as the ordinary coefficient of correlation computed on ranks, but its formula is simpler.

$$\rho = 1 - \frac{6\Sigma D_i^2}{n(n^2-1)}$$

Here, n is the number of observations and D_i, the positive difference between ranks associated with the individuals i.

Like r, the rank correlation lies between –1 and +1.

Example 1.38: The ranks given by two judges to 10 individuals are as follows:

Individual	*Rank given by* Judge I x	Judge II y	$D = x - y$	D^2
1	1	7	6	36
2	2	5	3	9
3	7	8	1	1
4	9	10	1	1
5	8	9	1	1
6	6	4	2	4
7	4	1	3	9
8	3	6	3	9
9	10	3	7	49
10	5	2	3	9
				$\Sigma D^2 = 128$

Solution: The rank correlation is given by,

$$\rho = 1 - \frac{6\Sigma D^2}{n^3 - n} = 1 - \frac{6 \times 128}{10^3 - 10} = 1 - 0.776 = 0.224$$

The value of $\rho = 0.224$ shows that the agreement between the judges is not high.

NOTES

Example 1.39: Consider example 1.38 and compute r and compare.

Solution: The simple coefficient of correlation r for the previous data is calculated as follows:

x	y	x^2	y^2	xy
1	7	1	49	7
2	5	4	25	10
7	8	49	64	56
9	10	81	100	90
8	9	64	81	72
6	4	36	16	24
4	1	16	1	4
3	6	9	36	18
10	3	100	9	30
5	2	25	4	10
$\Sigma x = 55$	$\Sigma y = 55$	$\Sigma x^2 = 385$	$\Sigma y^2 = 385$	$\Sigma xy = 321$

$$r = \frac{321-10\times\frac{55}{10}\times\frac{55}{10}}{\sqrt{385-10\times\left(\frac{55}{10}\right)^2}\sqrt{385-10\times\left(\frac{55}{10}\right)^2}} = \frac{18.5}{\sqrt{82.5\times82.5}} = \frac{18.5}{82.5} =$$

0.224

This shows that the Spearman ρ for any two sets of ranks is the same as the Pearson r for the set of ranks. But it is much easier to compute ρ.

Often, the ranks are not given. Instead, the numerical values of observations are given. In such a case, we must attach the ranks to these values to calculate ρ.

Example 1.40: On the basis of given table define correlation and calculate rank.

Marks in Maths	*Marks in Stats*	*Rank in Maths*	*Rank in Stats*	D	D^2
45	60	4	2	2	4
47	61	3	1	2	4
60	58	1	3	2	4
38	48	5	4	1	1
50	46	2	5	3	9
					$\Sigma D^2 = 22$

$$\rho = 1-\frac{6\Sigma D^2}{n^3-n} = 1-\frac{6\times22}{125-5} = -0.1$$

Solution: This shows a negative, though small, correlation between the ranks.

If two or more observations have the same value, their ranks are equal and obtained by calculating the means of the various ranks.

If in this data, marks in maths are 45 for each of the first two students, the rank of each would be $\frac{3+4}{2} = 3.5$. Similarly, if the marks of each of the last two students in statistics are 48, their ranks would be $\frac{4+5}{2} = 4.5$

The problem takes the following shape:

NOTES

Marks in Maths	Marks in Stats	Rank x	Rank y	D	D^2
45	60	3.5	2	1.5	2.25
45	61	3.5	1	2.5	6.25
60	58	1	3	2	4.00
38	48	5	4.5	0.5	0.25
50	48	2	4.5	2.5	6.25
				ΣD^2	= 19

$$\rho = 1 - \frac{6\Sigma D^2}{n^3 - n} = 1 - \frac{6 \times 19}{120} = +0.05$$

An elaborate formula which can be used in case of equal ranks is:

$$\rho = 1 - \frac{6}{n^3 - n}\left[\Sigma D^2 + \frac{1}{12}\Sigma(m^3 - m)\right].$$

Here, $\frac{1}{12}\Sigma(m^3 - m)$ is to be added to ΣD^2 for each group of equal ranks, m being the number of equal ranks each time.

For the given data, we have:

(*i*) For series x, the number of equal ranks $m = 2$.

(*ii*) For series y, also, $m = 2$; so that,

$$\rho = 1 - \frac{6}{5^3 - 5}\left[21 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]$$

$$= 1 - \frac{6}{120}\left[21 + \frac{6}{12} + \frac{6}{12}\right]$$

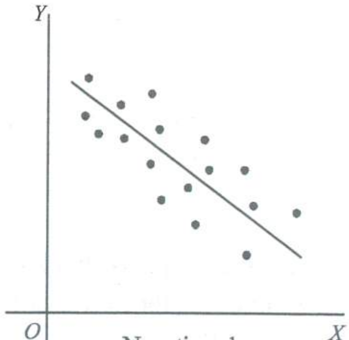
$$= 1 - \frac{6 \times 22}{120} = -0.1$$

Example 1.41: Show by means of diagrams various cases of scatter expressing correlation between x, y.

NOTES

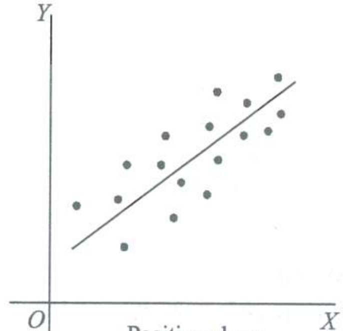
Solution:

(*i*)



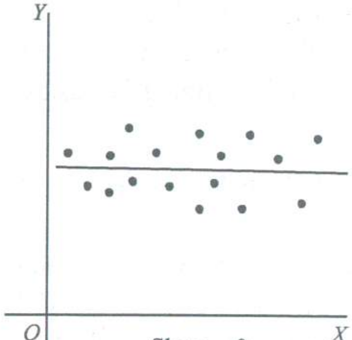
Negative slope
Inverse linear relationship
High scatter r low, negative

(*ii*)



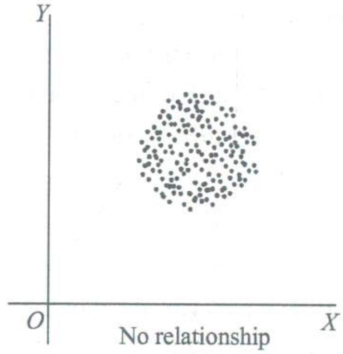
Positive slope
Direct linear relationship
High scatter r low, positive

(*iii*)



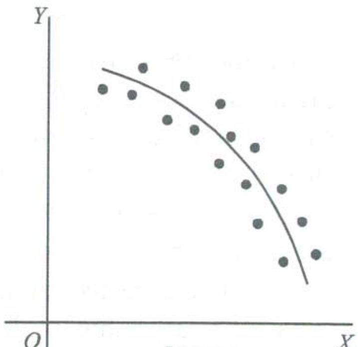
Slope = 0
$r = 0$

(*iv*)



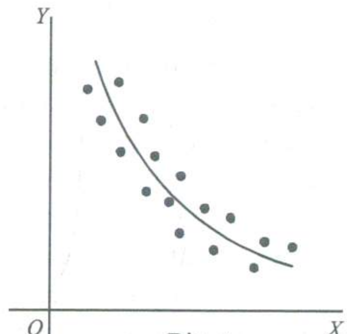
No relationship
$r \simeq 0$

(*v*)



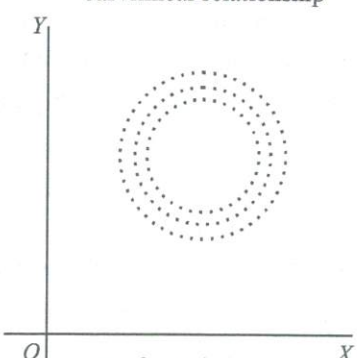
Inverse
curvilinear relationship

(*vi*)



Direct
curvilinear relationship

(*vii*)



Perfect relationship
But, $r = 0$ because of
non-linear relation

NOTES

Correlation analysis helps us in determining the degree to which two or more variables are related to each other.

When there are only two variables, we can determine the degree to which one variable is linearly related to the other. Regression analysis helps in determining the pattern of relationship between one or more independent variables and a dependent variable. This is done by an equation estimated with the help of data.

Uses of Rank Correlation

Correlation has many uses and definitions. Correlation helps to check how the two variables are related to each other. The two variables are termed as correlated when the change in one variable brings a change in the other variable.

A nonparametric or distribution-free rank statistic was introduced by Spearman for measuring the strength of the associations between two variables. Spearman rank correlation coefficient is used to define the *R*-estimate which is a measure of monotone association. It is specifically the uncertainty that interprets the significance of the linear correlation coefficient *r* for defining the basic concepts of *nonparametric* or *rank correlation*.

It is a statistical measure used to correlate nonparametric statistics when the data are in ordinal form.

Spearman's rank correlation coefficient evaluates how well the relationship between two variables can be described using a monotonic function. If there are no repeated data values, a perfect Spearman correlation of +1 or "1 occurs when each of the variables is a perfect monotone function of the other.

Spearman Rank Correlation Coefficient uses ranks to calculate correlation.

The sign of the Spearman correlation indicates the direction of association between *X* (the independent variable) and *Y* (the dependent variable). If *Y* tends to increase when *X* increases, the Spearman correlation coefficient is positive. If *Y* tends to decrease when *X* increases, the Spearman correlation coefficient is negative. A Spearman correlation of zero indicates that there is no tendency for *Y* to either increase or decrease when *X* increases. The Spearman correlation increases in magnitude as *X* and *Y* become closer to being perfect monotone functions of each other. A perfect monotone decreasing relationship implies that these differences always have opposite signs.

It also helps in calculating probability using a permutation test. An advantage of this approach is that it automatically takes into account the number of tied data values there are in the sample, and the way they are treated in computing the rank correlation.

A generalization of the Spearman coefficient is useful in the situation where there are three or more conditions, a number of subjects are all observed in each of them, and it is predicted that the observations will have a particular order.

It also measures correlation as parametric which are based on possible relationship of a parameterized form, such as a linear relationship.

It measures the data irrespective of the fact how the data is ranked, ascending or descending. Rank 1 can be assigned to the smallest value or the largest value as per the data analyst, provided the same ranking is done for both sets of data. The two sets of data are placed in rank order next to each other so that they can be accurately compared statistically.

(iv) Regression Equations

The term 'regression' was first used in 1877 by Sir Francis Galton who made a study that showed that the height of children born to tall parents will tend to move back or 'regress' toward the mean height of the population. He designated the word regression as the name of the process of predicting one variable from the another variable. He coined the term multiple regression to describe the process by which several variables are used to predict another. Thus, when there is a well established relationship between variables, it is possible to make use of this relationship in making estimates and to forecast the value of one variable (the unknown or the dependent variable) on the basis of the other variable/s (the known or the independent variable/s). A banker, for example, could predict deposits on the basis of per capita income in the trading area of bank. A marketing manager, may plan his advertising expenditures on the basis of the expected effect on total sales revenue of a change in the level of advertising expenditure. Similarly, a hospital superintendent could project his need for beds on the basis of total population. Such predictions may be made by using regression analysis. An investigator may employ regression analysis to test his theory having the cause and effect relationship. All this explains that regression analysis is an extremely useful tool specially in problems of business and industry involving predictions.

Assumptions in Regression Analysis

While making use of the regression techniques for making predictions, it is always assumed that:

(a) There is an actual relationship between the dependent and independent variables.

(b) The values of the dependent variable are random but the values of the independent variable are fixed quantities without error and are chosen by the experimentor.

(c) There is clear indication of direction of the relationship. This means that dependent variable is a function of independent variable. (For example, when we say that advertising has an effect on sales, then we are saying that sales has an effect on advertising).

(d) The conditions (that existed when the relationship between the dependent and independent variable was estimated by the regression) are the same when the regression model is being used. In other words, it simply means that the relationship has not changed since the regression equation was computed.

(e) The analysis is to be used to predict values within the range (and not for values outside the range) for which it is valid.

NOTES

(v) Simple Linear Regression Model

In case of simple linear regression analysis, a single variable is used to predict another variable on the assumption of linear relationship (i.e., relationship of the type defined by $Y = a + bX$) between the given variables. The variable to be predicted is called the dependent variable and the variable on which the prediction is based is called the independent variable.

Simple linear regression model[3] (or the Regression Line) is stated as,

$$Y_i = a + bX_i + e_i$$

Where, Y_i is the dependent variable

X_i is the independent variable

e_i is unpredictable random element (usually called as residual or error term)

(a) a represent the Y-intercept, i.e., the intercept specifies the value of the dependent variable when the independent variable has a value of zero. (But this term has practical meaning only if a zero value for the independent variable is possible).

(b) b is a constant, indicating the slope of the regression line. Slope of the line indicates the amount of change in the value of the dependent variable for a unit change in the independent variable.

If the two constants (viz., a and b) are known, the accuracy of our prediction of Y (denoted by $\hat{Y}$ and read as Y-hat) depends on the magnitude of the values of e_i. If in the model, all the e_i tend to have very large values then the estimates will not be very good but if these values are relatively small, then the predicted values ($\hat{y}$) will tend to be close to the true values (Y_i).

Estimating the Intercept and Slope of the Regression Model (or Estimating the Regression Equation)

The two constants or the parameters viz., 'a' and 'b' in the regression model for the entire population or universe are generally unknown and as such are estimated from sample information. The following are the two methods used for estimation:

(a) Scatter diagram method

(b) Least squares method

(a) Scatter diagram method

This method makes use of the Scatter diagram also known as Dot diagram. *Scatter diagram*[4] is a diagram representing two series with the known variable, i.e., independent variable plotted on the X-axis and the variable to be estimated, i.e., dependent variable to be plotted on the Y-axis on a graph paper (refer Figure 1.10) to get the following information:

Income X (Hundreds of Rupees)	Consumption Expenditure Y (Hundreds of Rupees)
41	44
65	60
50	39
57	51
96	80
94	68
110	84
30	34
79	55
65	48

The scatter diagram by itself is not sufficient for predicting values of the dependent variable. Some formal expression of the relationship between the two variables is necessary for predictive purposes. For the purpose, one may simply take a ruler and draw a straight line through the points in the scatter diagram and this way can determine the intercept and the slope of the said line and then the line can be defined as $\hat{Y} = a + bX_i$, with the help of which we can predict Y for a given value of X. But there are shortcomings in this approach. For example, if five different persons draw such a straight line in the same scatter diagram, it is possible that there may be five different estimates of a and b, specially when the dots are more dispersed in the diagram. Hence, the estimates cannot be worked out only through this approach. A more systematic and statistical method is required to estimate the constants of the predictive equation. The least squares method is used to draw the best fit line.

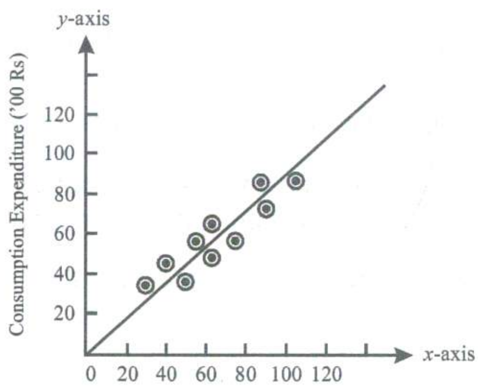


***Fig. 1.10** Scatter Diagram*

(b) Least squares method

Least squares method of fitting a line (the line of best fit or the regression line) through the scatter diagram is a method which minimizes the sum of the squared vertical deviations from the fitted line. In other words, the line to be fitted will pass through the points of the scatter diagram in such a way that the sum of the squares of the vertical deviations of these points from the line will be a minimum.

NOTES

The meaning of the least squares criterion can be easily understood through reference to Figure 1.11 drawn below, where Figure 1.10 in scatter diagram has been reproduced along with a line which represents the least squares line fit to the data.

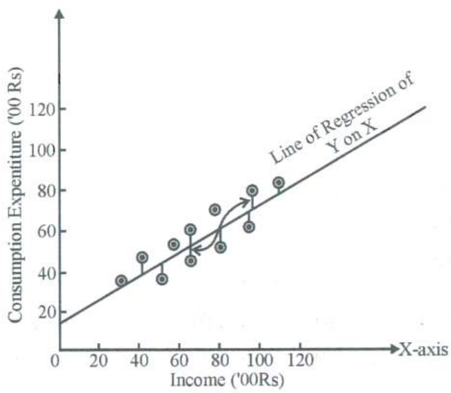


Fig. 1.11 Scatter Diagram, Regression Line and Short Vertical Lines Representing 'e'

Figure 1.10, the vertical deviations of the individual points from the line are shown as the short vertical lines joining the points to the least squares line. These deviations will be denoted by the symbol 'e'. The value of 'e' varies from one point to another. In some cases it is positive, while in others it is negative. If the line drawn happens to be least squares line, then the values of $\sum e_i$ is the least possible. It is so, because of this feature the method is known as Least Squares Method.

Why we insist on minimizing the sum of squared deviations is a question that needs explanation. If we denote the deviations from the actual value Y to the estimated value $\hat{Y}$ as $(Y-\hat{Y})$ or e_i, it is logical that we want the $\Sigma(Y-\hat{Y})$ or $\sum_{i=1}^{n} e_i$, to be as small as possible. However, mere examining $\Sigma(Y-\hat{Y})$ or $\sum_{i=1}^{n} e_i$, is inappropriate, since any e_i can be positive or negative. Large positive values and large negative values could cancel one another. But large values of e_i regardless of their sign, indicate a poor prediction. Even if we ignore the signs while working out $\sum_{i=1}^{n} |e_i|$, the difficulties may continue. Hence, the standard procedure is to eliminate the effect of signs by squaring each observation. Squaring each term accomplishes two purposes viz., (*i*) it magnifies (or penalizes) the larger errors, and (*ii*) it cancels the effect of the positive and negative values (since a negative error when squared becomes positive). The choice of minimizing the squared sum of errors rather than the sum of the absolute values implies that there are many small errors rather than a few large errors. Hence, in obtaining the regression line, we follow the approach that the sum of the squared deviations be minimum and on this basis work out the values of its constants viz., 'a' and 'b' also known as the intercept and the slope of the line. This is done with the help of the following two normal equations:

$$\Sigma Y = na + b\Sigma X$$

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

In the above two equations, 'a' and 'b' are unknowns and all other values viz., ΣX, ΣY, ΣX^2, ΣXY, are the sum of the products and cross products to be calculated from the sample data, and 'n' means the number of observations in the sample.

The following examples explain the Least squares method.

Example 1.42: Fit a regression line $\hat{Y} = a + bX_i$ by the method of Least squares to the given sample information.

Observations	1	2	3	4	5	6	7	8	9	10
Income (X) ('00 ₹)	41	65	50	57	96	94	110	30	79	65
Consumption Expenditure (Y) ('00 ₹)	44	60	39	51	80	68	84	34	55	48

Solution: We are to fit a regression line $\hat{Y} = a + bX_i$ to the given data by the method of Least squares. Accordingly, we work out the 'a' and 'b' values with the help of the normal equations as stated above and also for the purpose, work out ΣX, ΣY, ΣXY, ΣX^2 values from the given sample information table on Summations for Regression Equation.

Summations for Regression Equation

Observations	*Income X ('00 ₹)*	*Consumption Expenditure Y ('00 ₹)*	*XY*	*X^2*	*Y^2*
1	41	44	1804	1681	1936
2	65	60	3900	4225	3600
3	50	39	1950	2500	1521
4	57	51	2907	3249	2601
5	96	80	7680	9216	6400
6	94	68	6392	8836	4624
7	110	84	9240	12100	7056
8	30	34	1020	900	1156
9	79	55	4345	6241	3025
10	65	48	3120	4225	2304
$n = 10$	$\Sigma X = 687$	$\Sigma Y = 563$	$\Sigma XY = 42358$	$\Sigma X^2 = 53173$	$\Sigma Y^2 = 34223$

Putting the values in the required normal equations we have,

$$563 = 10a + 687b$$

$$42358 = 687a + 53173b$$

Solving these two equations for a and b we obtain,

$$a = 14.000 \quad \text{and} \quad b = 0.616$$

Hence, the equation for the required regression line is,

$$\hat{Y} = a + bX_i$$

or, $\hat{Y} = 14.000 + 0.616X_i$

This equation is known as the regression equation of Y on X from which Y values can be estimated for given values of X variable.

NOTES

Checking the accuracy of equation

After finding the regression line as stated above, one can check its accuracy also. The method to be used for the purpose follows from the mathematical property of a line fitted by the method of least squares viz., the individual positive and negative errors must sum to zero. In other words, using the estimating equation one must find out whether the term $\Sigma\left(Y - \hat{Y}\right)$ is zero and if this is so, then one can reasonably be sure that he has not committed any mistake in determining the estimating equation.

The problem of prediction

When we talk about prediction or estimation, we usually imply that if the relationship $Y_i = a + bX_i + e_i$ exists, then the regression equation, $\hat{Y} = a + bX_i$ provides a base for making estimates of the value for Y which will be associated with particular values of X. In Example 1.36, we worked out the regression equation for the income and consumption data as,

$$\hat{Y} = 14.000 + 0.616X_i$$

On the basis of this equation we can make a *point estimate* of Y for any given value of X. Suppose we wish to estimate the consumption expenditure of individuals with income of ₹ 10,000. We substitute $X = 100$ for the same in our equation and get an estimate of consumption expenditure as follows:

$$\hat{Y} = 14.000 + 0.616(100) = 75.60$$

Thus, the regression relationship indicates that individuals with ₹ 10,000 of income may be expected to spend approximately ₹ 7,560 on consumption. But this is only an expected or an estimated value and it is possible that actual consumption expenditure of same individual with that income may deviate from this amount and if so, then our estimate will be an error, the likelihood of which will be high if the estimate is applied to any one individual. The *interval estimate* method is considered better and it states an interval in which the expected consumption expenditure may fall. Remember that the wider the interval, the greater the level of confidence we can have, but the width of the interval (or what is technically known as the precision of the estimate) is associated with a specified level of confidence and is dependent on the variability (consumption expenditure in our case) found in the sample. This variability is measured by the standard deviation of the error term, '*e*', and is popularly known as the standard error of the estimate.

Standard error of the estimate

Standard error of estimate is a measure developed by the statisticians for measuring the reliability of the estimating equation. Like the standard deviation, the Standard

Error (S.E.) of $\hat{Y}$ measures the variability or scatter of the observed values of Y around the regression line. Standard Error of Estimate (S.E. of $\hat{Y}$) is worked out as under:

NOTES

$$\text{S.E. of } \hat{Y} \text{ (or } S_e) = \sqrt{\frac{\sum (Y-\hat{Y})^2}{n-2}} = \sqrt{\frac{\sum e^2}{n-2}}$$

where, S.E. of $\hat{Y}$ (or S_e) = Standard error of the estimate

Y = Observed value of Y

$\hat{Y}$ = Estimated value of Y

e = The error term = $(Y-\hat{Y})$

n = Number of observations in the sample

Note: In the above formula, $n-2$ is used instead of n because of the fact that two degrees of freedom are lost in basing the estimate on the variability of the sample observations about the line with two constants viz., 'a' and 'b' whose position is determined by those same sample observations.

The square of the S_e, also known as the variance of the error term, is the basic measure of reliability. The larger the variance, the more significant are the magnitudes of the e's and the less reliable is the regression analysis in predicting the data.

Interpreting the standard error of estimate and finding the confidence limits for the estimate in large and small samples

The larger the S.E. of estimate (SE_e), the greater happens to be the dispersion, or scattering, of given observations around the regression line. But if the S.E. of estimate happens to be zero then the estimating equation is a 'perfect' estimator (i.e., cent per cent correct estimator) of the dependent variable.

In case of large samples, i.e., where $n > 30$ in a sample, it is assumed that the observed points are normally distributed around the regression line and we may find,

68% of all points within $\hat{Y} \pm 1\ SE_e$ limits

95.5% of all points within $\hat{Y} \pm 2\ SE_e$ limits

99.7% of all points within $\hat{Y} \pm 3\ SE_e$ limits

This can be stated as,

(*i*) The observed values of Y are normally distributed around each estimated value of $\hat{Y}$ and;

(*ii*) The variance of the distributions around each possible value of $\hat{Y}$ is the same.

In case of small samples, i.e., where $n \leq 30$ in a sample the 't' distribution is used for finding the two limits more appropriately.

NOTES

This is done as follows:

$$\text{Upper limit} = \hat{Y} + `t' (SE_e)$$

$$\text{Lower limit} = \hat{Y} - `t' (SE_e)$$

Where, $\hat{Y}$ = The estimated value of Y for a given value of X.

SE_e = The standard error of estimate.

't' = Table value of 't' for given degrees of freedom for a specified confidence level.

Some other details concerning simple regression

Sometimes the estimating equation of Y also known as the regression equation of Y on X, is written as follows:

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

or,

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

Where, r = Coefficient of simple correlation between X and Y

σ_Y = Standard deviation of Y

σ_X = Standard deviation of X

$\overline{X}$ = Mean of X

$\overline{Y}$ = Mean of Y

$\hat{Y}$ = Value of Y to be estimated

X_i = Any given value of X for which Y is to be estimated.

This is based on the formula we have used, i.e., $\hat{Y} = a + bX_i$. The coefficient of X_i is defined as,

$$\text{Coefficient of } X_i = b = r\frac{\sigma_Y}{\sigma_X}$$

(Also known as regression coefficient of Y on X or slope of the regression line of Y on X) or b_{YX}.

$$= \frac{\Sigma XY - n\overline{XY} \times \sqrt{\Sigma Y^2 - n\overline{Y}^2}}{\sqrt{\Sigma Y^2 - n\overline{Y^2}}\sqrt{\Sigma X^2 - n\overline{X}^2}\sqrt{\Sigma X^2 - n\overline{X}^2}}$$

$$= \frac{\Sigma XY - n\overline{XY}}{\Sigma X^2 - n\overline{X}^2}$$

and

$$a = -r\frac{\sigma_Y}{\sigma_X}\overline{X} + \overline{Y}$$

$$= \overline{Y} - b\overline{X} \quad \left(\text{since } b = r\frac{\sigma_Y}{\sigma_X}\right)$$

Similarly, the estimating equation of *X*, also known as the regression equation of *X* on *Y*, can be stated as:

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

or

$$\hat{X} = r\frac{\sigma_X}{\sigma_Y}\left(Y - \overline{Y}\right) + \overline{X}$$

and the

Regression coefficient of *X* on *Y* (or b_{XY}) $= r\frac{\sigma_X}{\sigma_Y} = \frac{\sum XY - n\overline{XY}}{\sum Y^2 - n\overline{Y}^2}$

If we are given the two regression equations as stated above, along with the values of '*a*' and '*b*' constants to solve the same for finding the value of *X* and *Y*, then the values of *X* and *Y* so obtained, are the mean value of *X* (i.e., $\overline{X}$) and the mean value of *Y* (i.e., $\overline{Y}$).

If we are given the two regression coefficients (viz., b_{XY} and b_{YX}), then we can work out the value of coefficient of correlation by just taking the square root of the product of the regression coefficients as shown below:

$$r = \sqrt{b_{YX}.b_{XY}}$$

$$= \sqrt{r\frac{\sigma_Y}{\sigma_X}.r\frac{\sigma_X}{\sigma_Y}}$$

$$= \sqrt{r.r} \quad = r$$

The (±) sign of *r* will be determined on the basis of the sign of the regression coefficients given. If regression coefficients have minus sign then *r* will be taken with minus (–) sign and if regression coefficients have plus sign then *r* will be taken with plus (+) sign. (Remember that both regression coefficients will necessarily have the same sign whether it is minus or plus for their sign is governed by the sign of coefficient of correlation.)

Example 1.43: Given is the following information:

	$\overline{X}$	$\overline{Y}$
Mean	39.5	47.5
Standard Deviation	10.8	17.8

Simple correlation coefficient between *X* and *Y* is = + 0.42

Find the estimating equation of *Y* and *X*.

Solution:

Estimating equation of *Y* can be worked out as,

$$\because \quad \left(\hat{Y} - \overline{Y}\right) = r\frac{\sigma_Y}{\sigma_X}\left(X_i - \overline{X}\right)$$

or

$$\hat{Y} = r\frac{\sigma_Y}{\sigma_X}\left(X_i - \overline{X}\right) + \overline{Y}$$

NOTES

$$= 0.42\frac{17.8}{10.8}(X_i - 39.5) + 47.5$$

$$= 0.69X_i - 27.25 + 47.5$$

$$= 0.69X_i + 20.25$$

Similarly, the estimating equation of X can be worked out as under:

$$\because \quad (\hat{X} - \overline{X}) = r\frac{\sigma_X}{\sigma_Y}(Y_i - \overline{Y})$$

or $$\hat{X} = r\frac{\sigma_X}{\sigma_Y}(Y_i - \overline{Y}) + \overline{X}$$

or $$= 0.42\frac{10.8}{17.8}(Y_i - 47.5) + 39.5$$

$$= 0.26Y_i - 12.35 + 39.5$$

$$= 0.26Y_i + 27.15$$

Example 1.44: Given is the following data:

Variance of $X = 9$

Regression equations:

$$4X - 5Y + 33 = 0$$

$$20X - 9Y - 107 = 0$$

Find: (*i*) Mean values of X and Y.

(*ii*) Coefficient of Correlation between X and Y.

(*iii*) Standard deviation of Y.

Solution:

(*i*) For finding the mean values of X and Y, we solve the two given regression equations for the values of X and Y as follows:

$$4X - 5Y + 33 = 0 \quad (1)$$

$$20X - 9Y - 107 = 0 \quad (2)$$

If we multiply equation (1) by 5, we have the following equations:

$$20X - 25Y = -165 \quad (3)$$

$$20X - 9Y = 107$$

$$- \quad + \quad -$$

$$-16Y = -272$$

After subtracting equation (2) from (3)

or $$Y = 17$$

Putting this value of Y in equation (1) we have,

$$4X = -33 + 5(17)$$

or $$X = \frac{-33+85}{4} = \frac{52}{4} = 13$$

Hence, $$\overline{X} = 13 \quad \text{and} \quad \overline{Y} = 17$$

(*ii*) For finding the coefficient of correlation, first of all we presume one of the two given regression equations as the estimating equation of *X*. Let equation $4X - 5Y + 33 = 0$ be the estimating equation of *X*, then we have,

$$\hat{X} = \frac{5Y_i}{4} - \frac{33}{4}$$

and

From this we can write $b_{XY} = \frac{5}{4}$

The other given equation is then taken as the estimating equation of *Y* and can be written as,

$$\hat{Y} = \frac{20X_i}{9} - \frac{107}{9}$$

and from this we can write $b_{YX} = \frac{20}{9}$

If the above equations are correct then *r* must be equal to,

$$r = \sqrt{5/4 \times 20/9} = \sqrt{25/9} = 5/3 = 1.6$$

which is an impossible equation, since *r* can in no case be greater than 1. Hence, we change our supposition about the estimating equations and by reversing it, we re-write the estimating equations as under:

$$\hat{X} = \frac{9Y_i}{20} + \frac{107}{20}$$

and $$\hat{Y} = \frac{4X_i}{5} + \frac{33}{5}$$

Hence,
$$\begin{aligned} r &= \sqrt{9/20 \times 4/5} \\ &= \sqrt{9/25} \\ &= 3/5 \\ &= 0.6 \end{aligned}$$

Since, regression coefficients have plus signs, we take $r = + 0.6$

(*iii*) Standard deviation of *Y* can be calculated as follows:

$\because$ Variance of $X = 9$ $\quad\therefore$ Standard deviation of $X = 3$

$\because$ $$b_{YX} = r\frac{\sigma_Y}{\sigma_X} = \frac{4}{5} = 0.6\frac{\sigma_Y}{3} = 0.2\sigma_Y$$

Hence, $\sigma_Y = 4$

Alternatively, we can work it out as under:

$\because$ $$b_{XY} = r\frac{\sigma_X}{\sigma_Y} = \frac{9}{20} = 0.6\frac{\sigma_Y}{3} = \frac{1.8}{\sigma_Y}$$

Hence, $\sigma_Y = 4$

(vi) Forecasting

Forecasting involves the use of previous information for future predictions pertaining to style, systems, entrepreneurial techniques, etc. Mathematical principles are applied to locate the patterns and tests are conducted to examine the result for

NOTES

logic, thus this category of forecasting is termed as statistical forecasting. The process of statistical forecasting constitutes the following:

- **Multiple regression analysis:** This comes in use in cases where more than one independent factor is present. It is extensively used in intermediate term forecasting in order to determine the exclusion and inclusion of the factors. It may be used for the development of alternate models having varied factors.
- **Nonlinear regression:** This type of regression does not assume a linear relation amongst variables regularly used when time is the independent variable.
- **Trend analysis:** This method exercises linear and nonlinear regression with time as the explanatory variable used where pattern is over time.
- **Decomposition analysis:** This method identifies various patterns appearing simultaneously in a time series-time that consume each time it is used (also used to deseasonalize a series).
- **Moving average analysis:** These are simple moving averages, which are used to forecast future values according to a weighted average of past values that are easy to update.
- **Weighted moving averages:** These are quite useful and economical and are extensively applied in areas where repeatable forecasts are required. It constitutes the use of techniques such as, sum-of-the-digits and trend adjustment methods.
- **Adaptive filtering:** A type of moving average which includes a method of learning from past errors-can respond to changes in the relative importance of trend, seasonal, and random factors.
- **Exponential smoothing:** A moving average form of time series forecasting that is used with seasonal patterns and can be easily adjusted for previous mistakes. It is easy to prepare follow-on forecasts and is suited for conditions where multiple forecasts must be prepared. Various forms are used based on the trend or cyclical variations.
- **Hodrick-Prescott filter:** This is a smoothing technique for obtaining a long-term trend component in a time series. It is used for decomposing a given series into stationary and nonstationary components, such that the sum of squares of the series from the nonstationary component is a minimum, with a penalty on changes to the derivatives of the nonstationary component.
- **Modeling and simulation:** Modeling presents a situation through a series of equations allowing testing of the result of the changes in different factors. It is requires substantial consuming for construction and usually requires user programming or purchase of packages including SIMSCRIPT. It is very useful in developing and testing strategies otherwise non-evident.

Certainty models give only most likely outcome-advanced spreadsheets and may be used to conduct a 'what if' analysis, usually with the help of computer-based spreadsheets.

Probabilistic models utilize the Monte Carlo simulation techniques to handle uncertainty, giving a series of possible results for every set of events.

Forecasting error

Forecasting models have either an implicit or explicit error structure, where error is defined as the difference between the model prediction and the 'true' value. Further, various data snooping methods in the domain of statistics requires to be applied to data supplied to a forecasting model. Also, diagnostic checking, as defined within the field of statistics, is required for any model which uses data.

The application of any technique for forecasting requires the use of a performance measure to evaluate its quality. Mean Absolute Deviation (MAD), and variance are the best measures. However, MAD does not lend itself to further use making inferences, but that the standard error does. For the error analysis purposes variance is preferable as variances of independent (uncorrelated) errors are additive. MAD is not additive.

Check Your Progress

20. What are the limitations of geometric mean?
21. What is absolute measure of dispersion?
22. What is relative measure of dispersion?
23. Define range.
24. What do you understand by correlation analysis?
25. What does theory of correlation explain?
26. What phenomenon is necessary for correlation?
27. What are the measures of describing correlation?
28. Define coefficient of determinations.
29. What is coefficient correlation? How does it measure the degree of relationship?
30. Name the methods used to find coefficient of correlation.
31. What are the assumptions on which Karl Pearson coefficient is based?
32. What is rank correlation? Write its formula.
33. What is the meaning of regression analysis?
34. What is a simple linear regression model?
35. Define the two constants involved in regression.
36. Mention the methods to calculate the constants in regression models.
37. What is a scatter diagram method?
38. What is a least squares method?
39. What is the standard error of the estimate?

1.6 ANSWERS TO 'CHECK YOUR PROGRESS'

NOTES

1. Statistics originated from two quite dissimilar fields, viz., games of chance and political states. These two different fields are also termed as two distinct disciplines—one primarily analytical and the other essentially descriptive. The former is associated with the concept of chance and probability and the latter is concerned with the collection of data.
2. Statistical methods of analysis are helpful in the marketing function of an enterprise through its enormous help in market research, advertisement campaigns and in comparing the sales performances.
3. Under this method, the investigator presents himself personally before the informant and obtains a first-hand information. This method is most suitable when the field of enquiry is small and a greater degree of accuracy is required.
4. Under this method, the investigator prepares a questionnaire containing a number of questions pertaining to the field of enquiry. These questionnaires are sent by post to the informants together with a polite covering letter explaining in detail the aims and objectives of collecting the information and requesting the respondents to cooperate by furnishing the correct replies and returning the questionnaire duly filled in.
5. Following are the merits of interview method:
   (i) This method is less costly and less time-consuming than direct personal investigation.
   (ii) Under this method, the enquiry can be formulated and conducted more effectively and efficiently as it is possible to obtain the views and suggestions of the experts on the given problem.
6. Given below are the limitations of telephone survey method:
   (i) This method excludes those who do not have a telephone as also those who have unlisted telephones.
   (ii) This method is also subjective in nature and personal bias, whim and prejudices of the investigator may adversely affect the results of the enquiry.
7. Experiments are resorted to when it is required to collect factual data when nothing is available for reference. It may also be conducted to verify the theory. It is a study conducted under controlled conditions. Experiments are made by researchers to understand the cause and effect relationships. Such relationships are also made in observational studies but here, there is no control on how subjects are assigned to groups.
8. Tabulation is the systematic arrangement of data in columns and rows.
9. Tabulation has the following objectives:
   (i) Simplicity. The removal of unnecessary details gives a clear and concise picture of the data
   (ii) Economy of space and time
   (iii) Ease in comprehension and remembering

(iv) Facility of comparisons. Comparisons within a table and with other tables may be made

(v) Ease in handling of totals, analysis, interpretation, etc.

10. Here, the points are plotted on paper (or graph paper) and joined by straight lines. Generally, continuous variables are plotted by the line graph.
11. A frequency polygon is a line chart of frequency distribution in which, either the values of discrete variables or midpoints of class intervals are plotted against the frequencies and these plotted points are joined together by straight lines.
12. In a frequency distribution, if the frequency in each class interval is converted into a proportion, dividing it by the total frequency, we get a series of proportions called *relative frequencies*.
13. Cumulative frequency curve or ogive is the graphic representation of a cumulative frequency distribution.
14. Less than ogive: The less than cumulative frequencies are plotted against the upper boundaries of their respective class intervals.

    Greater than ogive: The greater than cumulative frequencies are plotted against the lower boundaries of their respective class intervals.
15. A histogram is the graphical description of data and is constructed from a frequency table. It displays the distribution method of a data set and is used for statistical as well as mathematical calculations.
16. Diagrams are more suitable to illustrate the data which is discrete, while continuous data is better represented by graphs.
17. Two dimensional diagrams take two components of data for representation. These are also called area diagrams as it considers two dimensions. The types are rectangles, squares and pie.
18. This type of diagram enables us to show the partitioning of a total into its component parts. The diagram is in the form of a circle and is also called a pie because the entire diagram looks like a pie and the components resemble slices cut from it.
19. Three dimensional diagrams are also termed as volume diagram and consist of cubes, cylinders, spheres, etc. In these diagrams, three dimensions namely length, width and height are taken into account. Cubes are used to represent where side of a cube is drawn in proportion to the cube root of the magnitude of data.
20. The major drawbacks of the geometric mean are: (i) It is difficult to use and to compute and (ii) It is determined for positive values and cannot be used for negative values of zero.
21. Absolute measure of dispersion states the actual amount by which an item on an average deviates from a measure of central tendency.
22. Relative measure of dispersion is a quotient computed by dividing the absolute measures by a quantity in respect to which absolute deviation has been computed.

23. The range of a set of numbers is the difference between the maximum and minimum values. It indicates the limits within which the values fall.
24. Correlation analysis is the statistical tool generally used to describe the degree to which, one variable is related to another.
25. The theory by means of which quantitative connections between two sets of phenomena are determined is called the 'Theory of Correlation'.
26. For correlation, it is essential that the two phenomena should have cause-effect relationship.
27. Statisticians have developed two measures for describing the correlation between two variables, viz., the coefficient of determination and the coefficient of correlation.
28. The coefficient of determination is a measure of the degree of linear association or correlation between two variables, say X and Y, one of which happens to be independent variable and the other being dependent variable.
29. The coefficient of correlation, symbolically denoted by 'r', is another important measure to describe how well one variable is explained by another. It measures the degree of relationship between the two casually related variables. The value of this coefficient can never be more than +1 or less than –1.
30. (*i*) Coefficient of Correlation by the Method of Least Squares.
    (*ii*) Coefficient of Correlation using Simple Regression Coefficients.
    (*iii*) Coefficient of Correlation through Product Moment Method or Karl Pearson's Coefficient of Correlation.
31. Karl Pearson coefficient is based on the following assumptions:
    (*i*) There is a linear relationship between the two variables which means that straight line would be obtained if the observed data are plotted on a graph.
    (*ii*) The two variables are casually related which means that one of the variables is independent and the other one is dependent.
    (*iii*) A large number of independent causes are operating in both the variables so as to produce a normal distribution.
32. The rank correlation, written r, is a descriptive index of agreement between ranks over individuals. It is the same as the ordinary coefficient of correlation computed on ranks, but its formula is simpler.
33. Regression analysis is an extremely useful tool especially in problems of business and industry for making predictions.
34. In simple linear regression model, a single variable is used to predict another variable on the assumption of linear relationship between the given variables. The variable to be predicted is called the dependent variable and the variable on which the prediction is based is called the independent variable.
35. The two constants involved in regression model are a and b, where a represents the Y-intercept and b indicates the slope of the regression line.

NOTES

36. There are two methods to calculate the constants in regression models. They are:
    (*i*) Scatter diagram method
    (*ii*) Least squares method
37. Scatter diagram is the method to calculate the constants in regression models that makes use of scatter diagram or dot diagram. A scatter diagram is a diagram that represents two series with the known variables, i.e., independent variable plotted on the *X*-axis and the variable to be estimated, i.e., dependent variable to be plotted on the *Y*-axis.
38. The least squares method is a method to calculate the constants in regression models for fitting a line through the scatter diagram that minimizes the sum of the squared vertical deviations from the fitted line. In other words, the line to be fitted will pass through the points of the scatter diagram in such a way that the sum of the squares of the vertical deviations of these points from line will be a minimum.
39. The standard error of estimate is a measure developed by statisticians for measuring the reliability of the estimating equation. Like the standard deviation, the standard error of estimate measures the variability or scatter of the observed values around the regression line.

1.7 SUMMARY

- Collection of facts is the first step in the statistical treatment of a problem. Numerical facts are the raw materials upon which the statistician is to work and just as in a manufacturing concern the quality of a finished product depends, inter alia, upon the quality of the raw material, in the same manner, the validity of statistical conclusions will be governed, among other considerations, by the quality of data used.
- The process of statistical analysis is a method of abstracting significant facts from the collected mass of numerical data. This process includes such things as 'measures of central tendency'—the determination, of Mean, Median and Mode—'measures of dispersion' and the determination of trends and tendencies, etc.
- Statistics originated from two quite dissimilar fields, viz., games of chance and political states. These two different fields are also termed as two distinct disciplines—one primarily analytical and the other essentially descriptive.
- Statistics deals with only those subjects of inquiry which are capable of being quantitatively measured and numerically expressed. This is an essential condition for the application of statistical methods.
- Statistical data is only approximately and not mathematically correct. Greater emphasis is being laid on the sampling technique of collecting data. This means that by observing, only a limited number of items we make an estimate of the characteristic of the entire population.

NOTES

- Statistical quality control is now being used in industry for establishing quality standards for products, for maintaining the requisite quality, and for assuring that the individual lots sold are of a given standard of acceptance.
- There are different methods of data collection. In the direct personal observation the investigator himself is present before the informant and obtains first-hand information.
- There are advantages and demerits in such observation. You have also learned what is a questionnaire, how is it drafted and what all things should be kept in mind while preparing it.
- There are two types of questionnaires with their own set of advantages and limitations.
- Interviews can also be indirect. Other types of data collection methods are telephone survey and information collected through local agents.
- Experiments are made by researchers to understand cause-and-effect relationships. Such relationships are also made in observational studies but here, there is no control on how subjects are assigned to groups.
- Besides collecting the appropriate data, a significant emphasis is also laid on the suitable representation of that data. This calls for using several data representation techniques, which depend on the nature and type of data collected.
- The data can be represented through frequency distributions, bar diagrams, pictograms, histograms and pie diagrams.
- The most important objective of statistical analysis is to determine a single value for the entire mass of data, so that it describes the overall level of the group of observations and can be considered a representative value of the whole set of data. It tells us where the centre of the distribution of data is located on the scale that we are using. This unit lays emphasis on the distribution of data and measures of central location.
- The geometric mean of n positive values is defined as the nth root of their product and is obtained by multiplying together all the values and then extracting the relevant root of the product.
- The measures of dispersion bring out this inequality. In engineering problems too the variability is an important concern.
- The amount of variability in dimensions of nominally identical components is critical in determining whether or not the components of a mass-produced item will be really interchangeable.
- The scatter in the life of a light bulb is at times more important than the average life. Thus, the measures of dispersion are useful in determining how representative the average is as a description of the data, comparing two or more series with regard to their scatter and designing a production control system which is based on the premise that if a process is under control, the variability it produces is the same over a period of time.

- Concepts of correlation analysis is the process of finding how accurately the line fits the observations, and if all the observations lie exactly on the line of best fit, the correlation is considered to be 1 or unity.
- The least squares method is the most widely used procedure for developing estimates of the model parameters.
- The correlation coefficient measures only the degree of linear association between two variables.
- Correlation analysis is the technique of studying how the variations in one series are related to variations in another series. Measurement of the degree of relationship between two or more variables is called correlation analysis.
- Regression analysis involves identifying the relationship between a dependent variable and one or more independent variables.
- A model of the relationship is hypothesized, and estimates of the parameter values are used to develop an estimated regression equation.
- Various tests are then employed to determine if the model is satisfactory. If the model is deemed satisfactory, the estimated regression equation can be used to predict the value of the dependent variable for the given values of independent variables.
- Regression analysis is the mathematical process of using observations to find the line of best fit through the data in order to make estimates and predictions about the behaviour of the variables. This line of best fit may be linear (straight) or curvilinear to some mathematical formula.

1.8 KEY TERMS

- **Statistics:** It refers to the numerical statements of facts in any department of inquiry placed in relation to each other.
- **Direct personal observation:** In direct personal observation, the investigator himself is present before the informant and obtains first-hand information.
- **Mailed questionnaire method:** In mailed questionnaire method, the investigator prepares a questionnaire containing a number of questions pertaining to the field of enquiry.
- **Questionnaire sent through enumerators:** In questionnaire sent through enumerators, the investigator appoints agents known as enumerators, who go to the respondents personally with the questionnaire and record the respondents' replies.
- **Indirect personal interviews:** In indirect personal interviews, the investigator interviews several third persons who are directly or indirectly concerned with the subject matter of the enquiry and who are in possession of the requisite information.
- **Telephone survey:** In telephone survey, the investigator contacts the informants on telephone and collects the information.

- **Table:** It is the systematic arrangement of data in columns and rows.
- **Frequency polygon:** It is a line chart of frequency distribution in which the values of discrete variables or midpoints of class intervals are plotted against the frequencies and these plotted points are joined together by straight lines.
- **Relative frequency:** It is the series of proportions achieved after converting each class interval into a proportion, dividing it by the total frequency.
- **Ogive curve:** It is a graphic representation of a cumulative frequency distribution.
- **Histogram:** It is the graphical description of data that is constructed from a frequency table. It displays the distribution method of a data set and is used for statistical as well as mathematical calculations.
- **Pie diagram:** It is a diagram that enables us to show the partitioning of a total into its component parts.
- **Mean:** It is an arithmetic average and measure of central location.
- **Mode:** It is a form of average that can be defined as the most frequently occurring value in the data.
- **Median:** It is a measure of central tendency that appears in the centre of an ordered data.
- **Quartile:** A quartile divides the data into four equal parts.
- **Decile:** A decile divides the total ordered data into ten equal parts.
- **Percentile:** A percentile divides the data into hundred equal parts.
- **Quartile deviation**: It is a type of range based on quartiles.
- **Mean deviation:** It is the mean deviation of a series of values is the arithmetic mean of their absolute deviations.
- **Standard deviation:** It is the square root of the average of the squared deviations from their mean of a set of observations.
- **Range**: Range is the difference between the maximum and minimum values of a set of number. It indicates the limits within which the values fall.
- **Correlation analysis:** It is a statistical tool which is used to describe the degree to which one variable is related to another.
- **Coefficient of determination:** It is a measure of the degree of linear association or correlation between two variables, one of which must be an independent variable and the other, a dependent variable.
- **Coefficient of correlation:** Symbolically denoted by 'r' it is an important measure to describe how well one variable is explained by another. It measures the degree of relationship between the two casually related variables.
- **Regression analysis:** It refers to the relationship used for making estimates and forecasts about the value of one variable (the unknown or the dependent variable) on the basis of the other variable/s (the known or the independent variable/s).

NOTES

- **Scatter diagram:** Scatter diagram, also known as a Dot diagram, is used to represent two series with the known variables, i.e., independent variable plotted on the *X*-axis and the variable to be estimated, i.e., dependent variable to be plotted on the *Y*-axis on a graph paper for the given information.
- **Standard error of estimate:** It is a measure developed by statisticians for measuring the reliability of the estimating equation.

1.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write in brief about the three stages that trace the development of statistics.
2. What is observation? Why it is important for data collection?
3. What are the advantages of the mailed questionnaire method?
4. State the limitations of questionnaires sent through enumerators.
5. Mention the features of indirect personal interview.
6. What are the limitations of the mailed questionnaire method?
7. What are the merits of questionnaires sent through enumerators?
8. What types of questions should a questionnaire contain?
9. State the types of tables.
10. How will you construct a frequency distribution?
11. Write any three advantages of tabulation.
12. What is relative frequency?
13. How is the data presented in a graphic form?
14. Differentiate between a pie chart and a histogram.
15. What is an arithmetic mean?
16. Differentiate between a mean and a mode.
17. Write three characteristics of mean.
18. Define quartiles, deciles and percentiles with suitable examples.
19. What is standard deviation? Why is it used in statistical evaluation of data?
20. What is variance?
21. What is the importance of correlation analysis?
22. How is the coefficient of determination found out?
23. State Karl Pearson's method of measuring coefficient of correlation.
24. What is the relationship between coefficient of non-determination and coefficient of alienation?
25. List the basic precautions and limitations of regression and correlation analyses.

NOTES

26. What is Spearman's rank correlation?
27. How will you predict the value of a dependent variable?
28. Differentiate between scatter diagram and least squares method.
29. How is the standard error of estimate calculated?

Long-Answer Questions

1. Explain the statistical procedure.
2. Describe the limitation of statistics.
3. Discuss some typical situations in business which can be analyzed using statistical techniques.
4. Explain various data collection methods.
5. Discuss the merits and demerits of both types of questionnaire.
6. Describe the points which must be considered while drafting a questionnaire.
7. Discuss the merits and demerits of interview method.
8. How is information received through local agents? Explain its merits and demerits.
9. What is experiment method? Explain its role in data collection.
10. Describe the essential features of a table.
11. Discuss the advantages of tabulation.
12. Marks obtained by 50 students in History paper of 100 marks are as follows:

78	25	25	35	32	31	34	41	44	43
44	20	48	39	45	45	36	35	48	47
36	50	31	42	35	68	68	72	37	15
60	20	47	47	53	38	49	37	51	61
34	76	79	23	18	72	65	42	62	45

(a) Arrange the data in a frequency distribution table in class-interval of length 5 units.

(b) Arrange the data in the form of a cumulative frequency table in class-interval of length 5 units.

(c) Arrange the data in the form of a relative frequency distribution table in the class-interval of length 5 units.

13. The following table gives the monthly income in rupees per family in a working class locality:

Monthly Earnings `	*No. of Families (f)*
600–700	10
700–800	15
800–900	20
900–1000	25
1000–1100	15
1100–1200	10
1200–1300	5

(*i*) Draw a historgram

(*ii*) Draw the two ogives

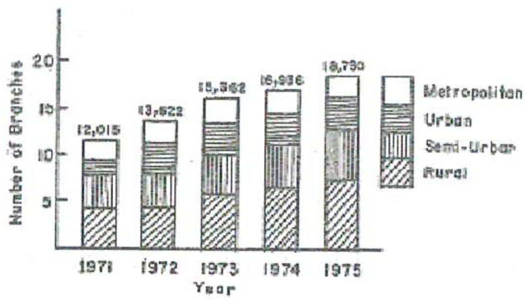
(*iii*) Draw a frequency polygon.

14. The following table gives the centre-wise distribution of the branches of commercial banks in India for the year 1971–75.

Year (30 *June Ending*)	Number of Branches				
	Rural	*Semi-Urban*	*Urban*	*Metropolitan*	*Total*
1971	4,279	4,016	1,776	1,940	12,015
1972	4,816	4,374	2,323	2,109	13,622
1973	5,561	4,723	2,573	2,505	15,362
1974	6,165	5,089	2,899	2,783	16,936
1975	6,806	5,569	3,267	3,088	18,730

Represent the data by means of a suitable diagram.

Centre-wise Distribution of Branches of Commercial Banks in India 1971–1975



15. Present the following data by means of a Pie diagram. Percentage of factors of Income Distribution for 1974–75.

	(*at current prices*)
1. Compensation of Employees	32.9%
2. Mixed Income of Self-Employed	53.9%
3. Interest	4.5%
4. Rent	3.3%
5. Profits and dividend	5.4%
Total	100.0%

16. The following table gives the heights (in inches) of 100 boys of a class. Calculate mean, mode and median of the height.

Height (inches)	*No. of Students*
60–62	5
62–64	18
64–66	42
66–68	20
68–70	8
70–72	7
	100

NOTES

17. The daily profits in rupees of 100 shops are distributed as follows. Draw a histogram of the data and then find the modal value. Check this value by direct calculation.

Profits per shop	*Number of shops*
0–100	12
100–200	18
200–300	27
300–400	20
400–500	17
500–600	6

Since class 200–300 has the highest frequency, i.e., 27, mode lies in this class.

Mode = 256.25

18. From the data given in the preceding question, draw a less than ogive curve, and locate median and quartiles. Check your value by direct calculations.

19. The following table gives the marks secured by 60 students of a class.

Marks	*Number of Students (f)*
10–20	8
20–30	12
30–40	20
40–50	10
50–60	7
60–70	3

Calculate mean, standard deviation and co-efficient of variation.

20. Two cricketers scored the following runs in ten innings. Find who is a better run getter and who is a more consistent player.

A: 42, 17, 83, 59, 72, 76, 64, 45, 40, 32

B: 28, 70, 31, 0, 59, 108, 82, 14, 3, 95

21. Calculate Karl Pearson's coefficient of Skewness for the following distribution.

Values	*Frequency*
5–10	5
10–15	10
15–20	15
20–25	20
25–30	15
30–35	10
35–40	5

22. Calculate coefficient of Quartile-deviation and Bowley's Quartile coefficient of skewness for the data given in Problem 6.

23. A researcher wants to find out if there is any relationship between the ages of the husbands and the ages of the wives. In other words, do old husbands

have old wives and young husbands have young wives? He took a random sample of 7 couples whose respective ages are given below:

Age of Husband (X)	*Age of Wife (Y)*
25	18
27	20
29	20
32	25
35	25
37	30
39	37

(a) For this data compute the regression line.

(b) Based upon the correlation between their ages, what would be the age of the wife, if the husband's age is 36 years.

24. On the basis of the data given in question 1, calculate the most probable age of husband if the age of wife is 28 years.

25. A researcher wants to find out if there is any relationship between the heights of the sons and the heights of the fathers. He took a random sample of six fathers and their six sons. Their heights in inches are given below in an ordered array:

Height of Father in inches (X)	*Height of Son in inches (Y)*
63	66
65	68
66	65
67	67
67	69
68	70

Using the short method

(i) Fit a regression line of Y on X, and hence predict the height of the son if father's height is 70 inches.

(ii) Fit a regression line of X on, Y and hence predict the height of the father is sons height is 65 inches.

(iii) Calculate karl Pearson's coefficient of correlation.

26. A department store gives in service training to its salesmen which is followed by a test. The following data give the test scores and sales made by nine salesmen during a certain period:

Test Scores :	14	19	24	21	26	22	15	20	19
Sales ('000 '):	31	36	48	37	50	45	33	41	39

Calculate Karl Pearson's coefficient of corelation between the test scores and sales.

27. The following data relate to advertising expenditure (in lakhs rupees) and sales (in crores of rupees) of a firm:

Advertising Expenditure :	10	12	15	23	20
(in lakh of ')					
Sales: (in crores of ')	14	17	23	25	21

(i) Estimate the sales target for an advertising expenditure of ' 25 lakhs.

(ii) Calculate the coefficient of correlation (r) and coefficient of determination of advertising expenditure and sales.

1.10 FURTHER READING

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House.

Gupta, S. C. 2006. *Fundamentals of Statistics*. New Delhi: Himalaya Publishing House.

Gupta, S. P., 2005. *Statistical Methods*. New Delhi: Sultan Chand and Sons.

Hooda, R. P. 2002. *Statistics for Business and Economics*. New Delhi: Macmillan India.

Kothari, C. R., 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Monga, G. S. 2000. *Mathematics and Statistics for Economics*. New Delhi: Vikas Publishing House

Gupta, S.P. 2006. *Statistical Methods*. New Delhi: S. Chand & Co. Ltd.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Gupta, S.C. and V.K. Kapoor. *Fundamentals of Mathematical Statistics*. New Delhi: Sultan Chand & Sons.

Sharma, A. 2009. *Operations Research.* New Delhi: Himalaya Publishing House.

Singh, J. K. 2009. *Business Mathematics*. New Delhi: Himalaya Publishing House.

Kalavathy, S. 2010. *Operations Research with C Programmes,* 3rd edition. New Delhi: Vikas Publishing House Pvt. Ltd.

Kothari, C. R. 1992. *An Introduction to Operational Research.* New Delhi: Vikas Publishing House Pvt. Ltd.

NOTES

UNIT 2 FORECASTING AND PROBABILITY CONCEPTS

Structure

2.0 INTRODUCTION

A time series is a set of ordered observations on a quantitative characteristic of a phenomenon at equally spaced time points. One of the main goals of time series analysis is to forecast future values of the series based on the trend which is a regular, slowly evolving change in the series level. Time series can be defined as a set of numeric observations of the dependent variable, measured at specific points in time in chronological order, usually at equal intervals in order to determine the relationship of time to such variable. You will learn that one of the major elements of planning and specifically strategic planning of any organization is accurately forecasting the future events that would have an impact on the growth of the organization. Previous performances must be studied to predict and forecast future activities. The quality of such forecasts is strongly related to the relevant information that can be extracted and used from past data. In that respect, time series can be used to determine the patterns in the data of the past over a period of time and extrapolate the data into the future. Thus, this method is very useful where the future planning is done after analysing the past data, methodologies and results. The underlying assumption in time series is that the same factors will continue to influence the future patterns of economic activity in a similar manner as they did in the past. Hence, the systematic analysis of time series graphs, and the types of variations that have influence on it, should be measured and treated to determine the pattern of future actions. This unit also deals with the concept of probability. The subject of probability in itself is a cumbersome one, hence only the basic concepts will be discussed in this unit. The word probability or chance is very commonly used in day-to-day conversation, and terms such as possible or probable or likely, all have similar meanings. Probability can be defined as a measure of the likelihood that a particular event will occur. The probability theory helps a decision-maker to analyse a situation and decide accordingly. Since the outcomes of most decisions cannot be accurately predicted because of the impact of many uncontrollable and unpredictable variables, it is necessary that all the known risks be scientifically evaluated. Probability theory, sometimes referred to as the science of uncertainty, is very helpful in such evaluations. It helps the decision-maker with only limited information to analyse the risks and select the strategy of minimum risk. This unit also discusses the laws of addition and multiplication. The law of addition states that when two events are mutually exclusive, the probability that either of the events will occur is the sum of their separate probabilities. The law of multiplication is applicable when two events occur at the same time. In this unit you will learn about probability distributions and Bayes' theorem. Bayes' theorem is a theorem of probability theory originally stated by the Reverend Thomas Bayes. It is based on the philosophy of science and tries to clarify the relationship between theory and evidence. In this unit, you will learn about the time series analysis, its importance in making forecasts and predictions and the concept of probability.

2.1 OBJECTIVES

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

- Discuss the influences of time series analysis
- Explain how smoothing techniques help in forecasting future trends
- Describe the techniques used in trend analysis
- Explain the method to measure cyclical effect
- State the steps involved in the calculation of seasonal variation
- Measure irregular variation and the seasonal adjustment of time series values
- Explain the basic concept of probability and various terms associated with it
- Describe marginal, conditional and joint probabilities
- Discuss addition theorem and multiplication theorem of probability
- State Bayes' theorem and its business applications

2.2 FORECASTING

The time series analysis method is quite accurate where the future is expected to be similar to the past. The underlying assumption in time series is that the same factors will continue to influence the future patterns of economic activity in a similar manner as in the past. These techniques are fairly sophisticated and require experts to use these methods.

NOTES

The classical approach to analysing a time series is in terms of four distinct types of variations or separate components that influence a time series. These components are:

(a) Secular Trend
(b) Cyclic Fluctuations
(c) Seasonal Variation
(d) Irregular Variation

(a) ***Secular Trend or Trend (T).*** The trend is the general long-term movement in the time series value of the variable (*Y*) over a fairly long period of time. The variable (*Y*) is the factor that we consider while making evaluation for the future. It could be sales, population, crime rate, and so on.

Trend is a common word, popularly used in day-to-day conversation, such as population trends, inflation trends, birth rate, and so on. These variables are observed over a long period of time and any changes related to time are noted and calculated and a trend of these changes is established. There are many types of trends; the series may be increasing slow or increasing fast or these may be decreasing at various rates. Some remain relatively constant and some reverse their trend from growth to decline or from decline to growth over a period of time. These changes occur as a result of a general tendency of the data to increase or decrease as a result of some identifiable influences.

If a trend can be determined and the rate of change can be ascertained, then tentative estimates on the same series of values into the future can be made. However, such forecasts are based upon the assumption that the conditions affecting the steady growth or decline are reasonably expected to remain unchanged in the future. A change in these conditions would affect the forecasts. A time series involving increase in population over time is illustrated graphically in Figure 2.1.

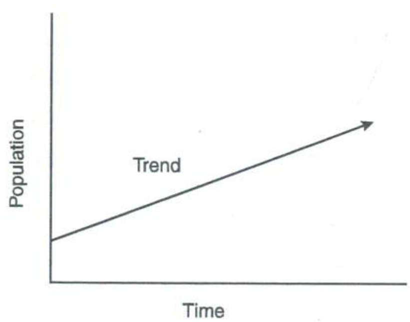


Fig. 2.1 *Time Series showing Increase in Population over Time*

(b) ***Cyclical Fluctuations (C).*** The cyclical fluctuations refer to regular swings or patterns that repeat over a long period of time. The movements are considered cyclical only if they occur after time intervals of more than one year. These are the changes that take place as a result of economic booms or depressions. These may be up or down, and are recurrent in nature and have a duration of several years, usually lasting for two to ten years. These movements also differ in

NOTES

intensity or amplitude and each phase of movement changes gradually into the phase that follows it. Some economists believe that the business cycle completes four phases every twelve to fifteen years, these four phases being: prosperity, recession, depression and recovery. However, there is no agreement on the nature or causes of these cycles.

Even though, measurement and prediction of cyclical variation is very important for strategic planning, the reliability of such measurements is highly questionable due to the following reasons:

(*i*) These cycles do not occur at regular intervals. In the twenty-five years from 1956 to 1981 in America, it is estimated that the peaks in the cyclical activity of the overall economy occurred in August 1957, April 1960, December 1969, November 1973 and January 1980. This shows that they differ widely in timing, intensity and pattern, thus making reliable evaluation of trends very difficult.

(*ii*) The cyclic variations are affected by many erratic, irregular and random forces which cannot be isolated and identified separately, nor can their impact be measured accurately.

Figure 2.2 shows graphically the cyclic variation, for revenues in an industry against time.

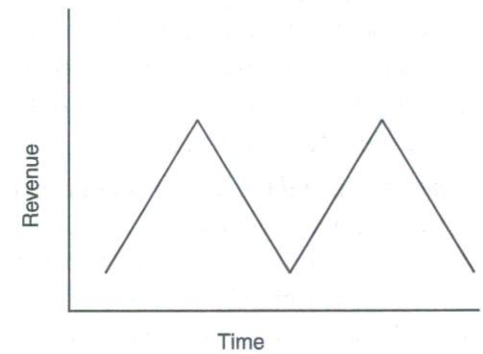


Fig. 2.2 *Cyclic Variation for Revenues aganist Time*

(c) *Seasonal Variation (S).* Seasonal variation involves patterns of change that repeat over a period of one year or less. Then they repeat from year to year and are brought about by fixed events. Sales of consumer items, for example, increase prior to Christmas due to the tradition of giving gifts. The sale of automobiles in America are much higher during the last three to four months of the year due to the introduction of new models. This data may be measured monthly or quarterly.

Since these variations repeat during a period of twelve months, they can be predicted fairly and accurately. Some factors that cause seasonal variations are:

(*i*) **Season and Climate.** Changes in the climate and weather conditions have a profound effect on sales. The sale of umbrellas in India, for example, is always more during monsoon or rainy season. Similarly, during winter, there is a greater demand for woollen clothes and hot drinks, while during the months of summer there is an increase in sales of fans and air conditioners.

(*ii*) **Customs and Festivals.** Customs and traditions affect the pattern of seasonal spending. Mother's Day or Valentine's Day in America, for example, see an increase in gift sales preceding these days. In India, festivals, such as Baisakhi and Diwali mean a big demand for sweets and candy. It is customary all over the world to give presents to children when they graduate from high school or college. Accordingly, the month of June, when most students graduate, is a time for the increase of sale for presents befitting the young.

(d) *Irregular (Random) Variation (I).* These variations are accidental, random or simply due to chance factors. Thus, they are wholly unpredictable. These fluctuations may be caused by such isolated incidents as floods, famines, strikes or wars. Sudden changes in demand or a breakthrough in a technological development may be included in this category. Accordingly, it is almost impossible to isolate and measure the value and the impact of these erratic movements on forecasting models or techniques. This phenomenon is graphically shown in Figure 2.3.

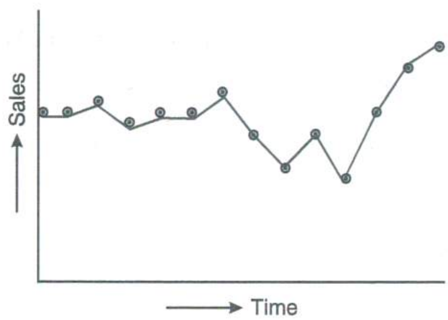


Fig. 2.3 Irregular Variation

It is traditionally acknowledged that the value of the time series (Y) is a function of the impact of variable trend (T), seasonal variation (S), cyclical variation (C) and irregular fluctuation (I). These relationships may vary depending upon assumptions and purposes. The effects of these four components might be additive, multiplicative, or a combination thereof in a number of ways. However, the traditional time series analysis model is characterized by multiplicative relationship, so that:

$$Y = T \times S \times C \times I$$

The above model is appropriate for those situations where percentage changes best represent the movement in the series and the components are not viewed as absolute values but as relative values.

Another approach to define the relationship may be additive, so that:

$$Y = T + S + C + I$$

This model is useful when the variations in the time series are in absolute values and can be separated and traced to each of these four parts and each part can be measured independently.

2.2.1 Smoothing Techniques

Smoothing techniques improve the forecasts of future trends provided that the time series are fairly stable with no significant trend, cyclical or seasonal effect and the objective is to *smooth out* the irregular component of the time series through the averaging process. There are two techniques that are generally employed for such smoothing.

NOTES

Moving Averages

The concept of the moving averages is based on the idea that any large irregular component of time series at any point in time will have a less significant impact on the trend, if the observation at that point in time is averaged with such values immediately before and after the observation under consideration. If, for example, we are interested in computing the three-period moving average for any time period, then we will take the average of the value in such time period, the value in the period immediately preceding it and the value in the time period immediately following it. Let us illustrate this concept by an example.

The table below represents the number of cars sold in the first 6 weeks of the first two months of the year by a given dealer. Our objective is to calculate a 3-week moving average.

Cars Sold in the First Two Months of the Year

Week	*Sales*
1	20
2	24
3	22
4	26
5	21
6	22

The moving average for the first 3-week period is given as:

$$\text{Moving average} = \frac{20+24+22}{3} = \frac{66}{3} = 22$$

This moving average can then be used to forecast the sale of cars for week 4. Since, the actual number of cars sold in week 4 is 26, we note that the error in the forecast is (26 – 22) = 4.

The calculation for the moving average for the next 3 periods is done by adding the value for week 4 and dropping the value for week 1, and taking the average for weeks 2, 3 and 4. Hence,

$$\text{Moving average} = \frac{24+22+26}{3} = \frac{72}{3} = 24$$

Then, this is considered to be the forecast of sales for week 5. Since the actual value of the sales for week 5 is 21, we have an error in our forecast of (21 – 24) = – (3).

The next moving average for weeks 3 to 5, as a forecast for week 6 is given as:

$$\text{Moving average} = \frac{22+26+21}{3} = \frac{69}{3} = 23$$

The error between the actual and the forecast value for week 6 is (22 – 23) = – (1). Since, the actual value of the sales for week 7 is not given, there is no need to forecast such values.

Our objective is to predict the trend and forecast the value of a given variable in the future as accurately as possible so that the forecast is reasonably free from random variations. To do that, we must have the sum of individual errors, as discussed above, as little as possible. However, since errors are irregular and random, it is expected that some errors would be positive in value and others negative, so that the sum of these errors would be highly distorted and would be closer to zero. This difficulty can be avoided by squaring each of the individual forecast errors and then taking the average. Naturally, the minimum values of these errors would also result in the minimum value of the 'average of the sum of squared errors'. This is shown as follows:

Week	*Time Series Value*	*Moving Average*	*Error*	*Error Squared*
1	20			
2	24			
3	22			
4	26	22	4	16
5	21	24	– 3	9
6	22	23	– 1	1

Then, the average of the sum of squared errors, also known as *mean squared error* and denoted by MSE, is given as:

$$\text{MSE} = \frac{16+9+1}{3} = \frac{26}{3} = 8.67$$

The value of MSE is an often-used measure of the accuracy of the forecasting method and the method which results in the least value of MSE is considered more accurate than others. The value of MSE can be manipulated by varying the number of data values to be included in the moving average. If, for example, we had calculated the value of MSE by taking 4 periods into consideration for calculating the moving average, rather than 3, then the value of MSE would be less. Accordingly, by using trial and error method, the number of data values selected for use in forecasting would be such that the resulting MSE value would be minimum.

Exponential Smoothing

In the moving average method, each observation in the moving average calculation receives the same weight. In other words, each value contributes equally towards the calculation of the moving average, irrespective of the number of time periods taken into consideration. In most actual situations, this is not a realistic assumption. As a result of the dynamics of the environment over a period of time, it is more likely that the forecast for the next period would be closer to the most recent previous period than the more distant previous period, so that the more recent value should get more weight than the previous value, and so on. The *exponential smoothing* technique uses the moving average with appropriate weights assigned to the values taken into consideration in order to arrive at a more accurate or smoothed forecast. It takes into consideration the decreasing impact of the past time periods as we move further into the past time periods. This decreasing impact as we move down into the time period is exponentially distributed and hence, the name *exponential smoothing*.

In this method, the smoothed value for period t, which is the weighted average of that period's actual value and the *smoothed* average from the previous period

NOTES

$(t-1)$, becomes the forecast for the next period $(t+1)$. Then, the exponential smoothing model for time period $(t+1)$ can be expressed as follows:

$$F_{(t+1)} = \alpha Y_t + (1-\alpha)F_t$$

Where, $F_{(t+1)}$ = The forecast of the time series for period $(t + 1)$

Y_t = Actual value of the time series in period t

α = Smoothing factor $(0 \leq \alpha \leq 1)$

F_t = Forecast of the time series for period t

The value of α is selected by the decision maker on the basis of the degree of smoothing required. A small value of α means a greater degree of smoothing. A large value of α means very little smoothing. When $\alpha = 1$, then there is no smoothing at all, so that the forecast for the next time period is exactly the same as the actual value of time series in the current period. This can be worked out as:

$$F_{(t+1)} = \alpha Y_t + (1-\alpha)F_t$$

When, $\alpha = 1$

$$F_{(t+1)} = Y_t + 0F_t = Y_t$$

The exponential smoothing approach is simple to use and once the value of α is selected, it requires only two pieces of information namely Y_t and F_t to calculate $F_{(t+1)}$.

To begin with the exponential smoothing process, let F_t be equal to the actual value of the time series in period t, which is Y_1. Hence, the forecast for period 2 is written as:

$$F_2 = \alpha Y_1 + (1-\alpha)\, F_1$$

But since we have put $F_1 = Y_1$, hence,

$$F_2 = \alpha Y_1 + (1-\alpha)Y_1$$
$$= Y_1$$

Let us now apply exponential smoothing method to the problem of forecasting car sales as discussed in the case of moving averages. The data once again is given as follows:

Week	*Time Series Value (Y_t)*
1	20
2	24
3	22
4	26
5	21
6	22

Let $\alpha = 0.4$

Since F_2 is calculated above as equal to $Y_1 = 20$, we can calculate the value of F_3 as follows:

$$F_3 = 0.4Y_2 + (1-0.4)\, F_2$$

Since, $F_2 = Y_1$, we get

$$F_3 = 0.4(24) + 0.6(20) = 9.6 + 12 \quad = 21.6$$

Similar values can be calculated for subsequent periods, so that:

NOTES

$$F_4 = 0.4Y_3 + 0.6F_3$$
$$= 0.4(22) + 0.6(21.6)$$
$$= 8.8 + 12.96 \quad = 21.76$$
$$F_5 = 0.4Y_4 + 0.6F_4$$
$$= 0.4(26) + 0.6(21.76)$$
$$= 10.4 + 13.056 \quad = 23.456$$
$$F_6 = 0.4Y_5 + 0.6F_5$$
$$= 0.4(21) + 0.6(23.456)$$
$$= 8.4 + 14.07 \quad = 22.47$$

and
$$F_7 = 0.4Y_6 + 0.6F_6$$
$$= 0.4(22) + 0.6(22.47)$$
$$= 8.8 + 13.48 \quad = 22.28$$

Now we can compare the exponential smoothing forecast value with the actual values for the six time periods and calculate the forecast error.

Week	*Time Series Value* (Y_t)	*Exponential Smoothing Forecast Value* (F_t)	*Error* $(Y_t - F_t)$
1	20	–	–
2	24	20.000	4.0
3	22	21.600	0.4
4	26	21.760	4.24
5	21	23.456	−2.456
6	22	22.470	−0.47

(The value of F_7 is not considered because the value of Y_7 is not given).

Let us now calculate the value of MSE for this method with selected value of $\alpha = 0.4$. From the previous table:

Forecast Error $(Y_t - F_t)$	*Squared Forecast Error* $(Y_t - F_t)^2$
0.4	0.16
4.24	17.98
−2.456	6.03
−0.47	0.22
	Total = 40.39

Then,

$$\text{MSE} = 40.39/5$$
$$= 8.08$$

The previous value of MSE was 8.67. Hence, the current approach is a better one.

The choice of the value for α is very significant. Let us look at the exponential smoothing model again.

NOTES

$$F_{(t+1)} = \alpha Y_t + (1-\alpha)F_t$$

$$= \alpha Y_t + F_t - \alpha F_t \quad = F_t + \alpha(Y_t - F_t)$$

where $(Y_t - F_t)$ is the forecast error in time period t.

The accuracy of the forecast can be improved by carefully selecting the value of α. If the time series contains substantial random variability then a small value of α (known as smoothing factor or smoothing constant) is preferable. On the other hand, a larger value of α would be desirable for time series with relatively little random variability $(Y_t - F_t)$.

2.2.2 Trend Analysis

While the chance variations are difficult to identify, separate, control or predict, a more precise measurement of trend, cyclical effects and seasonal effects can be made in order to make the forecasts more reliable.

When a time series shows an upward or downward long-term linear trend, then regression analysis can be used to estimate this trend and project the trends into forecasting the future values of the variables involved. Regression analysis has already been discussed extensively in Unit 1, where we described as to how the method of least squares could be used to find the best straight line relationship (line of best fit) between two variables. The equation for the straight line we used to describe the linear relationship between independent variable X and dependent variable Y was:

$$\hat{Y} = a + bX_i$$

Where, a = Intercept on the Y-axis, and b = Slope of the straight line.

In time series analysis the independent variable is time, so we will use the symbol t in place of X and we will use the symbol Y_t in place of $\hat{Y}$ which we have used previously.

Hence, the equation for linear trend is given as:

$$Y_t = a + bt$$

Where, Y_t = Forecast value of the time series in time period t

a = Intercept of the trend line on Y-axis

b = Slope of the trend line

t = Time period

As discussed earlier, we can calculate the values of a and b by the following formulae:

$$b = \frac{n\Sigma(tY) - (\Sigma t)(\Sigma Y)}{n(\Sigma t^2) - (\Sigma t)^2}, \text{ and } a = \overline{Y} - b\overline{t}$$

Where, Y = Actual value of the time series in time period t

n = Number of periods

$\overline{Y}$ = Average value of time series $= \frac{\Sigma Y}{n}$

$\overline{t}$ = Average value of $t = \frac{\Sigma t}{n}$

Knowing these values, we can calculate the value of Y_t.

NOTES

Example 2.1: A car fleet owner has 5 cars which have been in the fleet for several different years. The manager wants to establish if there is a linear relationship between the age of the car and the repairs in hundreds of dollars for a given year. This way, he can predict the repair expenses for each year as the cars become older. The information for the repair costs he collected for the last year on these cars is given below:

Car #	*Age (t)*	*Repairs (Y)*
1	1	4
2	3	6
3	3	7
4	5	7
5	6	9

The manager wants to predict the repair expenses for next year for the two cars that are 3 years old now.

Solution: The trend in repair costs suggests a linear relationship with the age of the car, so that the linear regression equation is given as:

$$Y_t = a + bt$$

Where, $$b = \frac{n\Sigma(tY) - (\Sigma t)(\Sigma Y)}{n(\Sigma t^2) - (\Sigma t)^2} \quad \text{and} \quad a = \overline{Y} - b\overline{t}$$

To calculate the various values, let us form a new table as follows:

Age of Car (t)	*Repair Cost (Y)*	*tY*	$\overline{t}^2$
1	4	4	1
3	6	18	9
3	7	21	9
5	7	35	25
6	9	54	36
Total 18	33	132	80

Knowing that $n = 5$, let us substitute these values to calculate the regression coefficients a and b.

Then, $$b = \frac{5(132) - (18)(33)}{5(80) - (18)^2} = \frac{660 - 594}{400 - 324} = \frac{66}{76} = 0.87$$

and $$a = \overline{Y} - b\overline{t}$$

Where, $$\overline{Y} = \frac{\Sigma Y}{n} = \frac{33}{5} = 6.6 \quad \text{and} \quad \overline{t} = \frac{t}{n} = \frac{18}{5} = 3.6$$

Then, $$a = 6.6 - 0.87(3.6)$$
$$= 6.6 - 3.13 \quad = 3.47$$

Hence, $$Y_t = 3.47 + 0.87t$$

The cars that are 3 years old now will be 4 years old next year, so that $t = 4$.

Hence, $$Y_{(4)} = 3.47 + 0.87(4)$$
$$= 3.47 + 3.48 \quad = 6.95$$

NOTES

Accordingly, the repair costs on each car that is 3 years old now are expected to be $695.00.

2.2.3 Measuring the Cyclical Effect

Cyclic variation is a pattern that repeats over time periods longer than one year. These variations are generally unpredictable in relation to the time of occurrence, duration as well as amplitude. However, these variations have to be separated and identified. The measure we use to identify cyclical variation is the *percentage of trend* and the procedure used is known as the *residual trend.*

As we have discussed before, there are four components of time series. These are secular trend (T), seasonal variation (S), cyclical variation (C) and irregular (or chance) variation (I). Since the time period considered for seasonal variation is less than one year, it can be excluded from the study because, when we look at time series consisting of annual data spread over many years, then only the secular trend, cyclical variation and irregular variation are considered.

Since secular trend component can be described by the trend line (usually calculated by line of regression), we can isolate cyclical and irregular components from the trend. Furthermore, since irregular variation occurs by chance and cannot be predicted or identified accurately, it can be reasonably assumed that most of the variation in time series left unexplained by the trend component can be explained by the cyclical component. In that respect, cyclical variation can be considered as the *residual*, once other causes of variation have been identified.

The measure of cyclic variation as *percentage of trend* is calculated as follows:

(1) Determine the trend line (usually by regression analysis)

(2) Compute the trend value Y_t for each time period (t) under consideration

(3) Calculate the ratio Y/Y_t for each time period

(4) Multiply this ratio by 100 to get the percentage of trend, so that:

$$\text{Percentage of Trend} = \left(\frac{Y}{Y_t}\right)100$$

Example 2.2: The following is the data for energy consumption (measured in quadrillions of BTU) in the United States from 1981 to 1986 as reported in the Statistical Abstracts of the United States.

Year	*Time Period (t)*	*Annual Energy Consumption (Y)*
1981	1	74.0
1982	2	70.8
1983	3	70.5
1984	4	74.1
1985	5	74.0
1986	6	73.9

Assuming a linear trend, calculate the percentage of trend for each year (cyclical variation).

Solution: First we find the secular trend by the regression line method, which is given by:

$$Y_t = a + bt$$

NOTES

Where, $b = \dfrac{n\Sigma(tY) - (\Sigma t)(\Sigma Y)}{n(\Sigma t^2) - (\Sigma t)^2}$ and $a = \overline{Y} - b\overline{t}$

Let us make a table for these values.

t	Y	tY	t^2
1	74.0	74.0	1
2	70.8	141.6	4
3	70.5	211.5	9
4	74.1	296.4	16
5	74.0	370.0	25
6	73.9	443.4	36
$\Sigma t = 21$	$\Sigma Y = 437.3$	$\Sigma tY = 1536.9$	$\Sigma t^2 = 91$

Substituting these values we get,

$$b = \frac{6(1536.9) - (21)(437.3)}{6(91) - (21)^2}$$

$$= \frac{9221.4 - 9183.3}{546 - 441} \quad = \frac{38.1}{105} = 0.363$$

and $a = \overline{Y} - b\overline{t}$

Where, $\overline{Y} = \dfrac{\Sigma Y}{n} = \dfrac{437.3}{6} = 72.88 \qquad \overline{t} = \dfrac{21}{6} = 3.5$

Hence, $a = 72.88 - 0.363(3.5)$

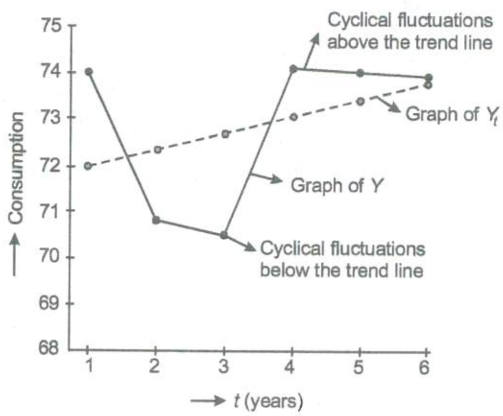
$= 72.88 - 1.27 \quad = 71.61$

Then, $Y_t = 71.61 + 0.363t$

Calculating the value of Y_t for each time period, we get the following table for percentage of trend $(Y/Y_t)100$.

Time Period (t)	*Energy Consumption* (Y)	*Trend Line* (Y_t)	*Percentage of Trend* $(Y/Y_t)100$
1	74.0	71.97	102.82
2	70.8	72.34	97.87
3	70.5	72.70	96.97
4	74.1	73.06	101.42
5	74.0	73.43	100.77
6	73.9	73.79	100.15

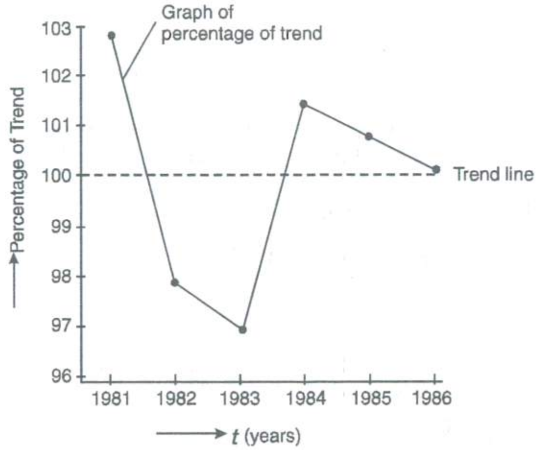
The following graph shows the actual energy consumption (Y), trend line (Y_t) and the cyclical fluctuations above and below the trend line over the time period (t) for 6 years.



NOTES

Frequently, we draw a graph of cyclic variation as the percentage of trend. This process eliminates the trend line and isolates the cyclical component of the time series.

It must be understood that cyclical fluctuations are not accurately predictable, and hence we cannot predict the future cyclic variations based upon such past cyclic variations.



The percentage of trend figures shows that in 1981, the actual consumption of energy was 102.82% of expected consumption that year and in 1983, the actual consumption was 96.97% of the expected consumption.

2.2.4 Seasonal Variation

Seasonal variation has been defined as predictable and repetitive movement around the trend line in a period of one year or less. For the measurement of seasonal variation, the time interval involved may be in terms of days, weeks, months or quarters. Since seasonal trends are predictable, we can plan in advance to meet these variations. Study of seasonal variations in the production data, for example, makes it possible to plan for hiring of additional personnel for peak periods of production or to accumulate an inventory of raw materials or to allocate vacation time to personnel, and so on.

In order to isolate and identify seasonal variations, we first eliminate, as far as possible, the effects of trend, cyclical variations and irregular fluctuations on the time series. Some of the methods used for the measurement of seasonal variations are described below:

Simple Average Method

This is the simplest method of isolating seasonal fluctuations in time series. It is based on the assumption that the series contain only the seasonal and irregular fluctuations. Assume that the time series involve monthly data over a time period of 5 years. Assume further that we want to find the seasonal index for the month of March (The seasonal variation will be the same for March in every year. Seasonal index describes the degree of seasonal variation.)

Then, the seasonal index for the month of March will be calculated as follows:

$$\text{Seasonal Index for March} = \left(\frac{\text{Monthly average for March}}{\text{Average of monthly averages}}\right) \times 100$$

NOTES

The following steps can be used in the calculation of seasonal index (variation) for the month of March (or any month), over the 5-year-period, regarding the sale of cars by one distributor.

1. Calculate the average sale of cars for the month of March over the last 5 years.

2. Calculate the average sale of cars for each month over the 5 years and then calculate the average of these monthly averages.

3. Use the above formula to calculate the seasonal index for March.

Let us say that the average sale of cars for the month of March over the period of 5 years is 360, and the average of all monthly average is 316. Then, the seasonal index for March = (360/316)×100 = 113.92.

Ratio to Moving Average Method

This is the most widely-used method of measuring seasonal variations. The seasonal index is based upon a mean of 100 with the degree of seasonal variation (seasonal index) measured by variations away from this base value. If, for example, we look at the seasonality of rental of row boats at a lake during the three summer months (a quarter) and we find that the seasonal index is 135 and we also know that the total boat rentals for the entire last year was 1680, then we can estimate the number of summer rentals for the row boats.

The average number of quarterly boats rented = 1680/4 = 420

The seasonal index, 135, for the summer quarter means that the summer rentals are 135 per cent of the average quarterly rentals.

Hence, summer rentals = 420 × (135/100) = 567

The steps required to compute the seasonal index can be enumerated by illustrating an example.

Example 2.3: Assume that a record of rental of row boats for the last 3 years on a quarterly basis is given as follows:

Year		*Rentals Per Quarter*		*Total*	
	I	II	III	IV	
1991	350	300	450	400	1500
1992	330	360	500	410	1600
1993	370	350	520	440	1680

Step 1. The first step is to calculate the four-quarter moving total for time series. This total is associated with the middle data point in the set of values for the four quarters as shown below.

Year	*Quarters*	*Rentals*	*Moving Total*
1991	I	350	
	II	300	
			1500
	III	450	
	IV	400	

The moving total for the given values for four quarters is 1500 which is simply the addition of the values of four quarters. This value of 1500 is placed in the

NOTES

middle of values 300 and 450 and recorded in the next column. For the next moving total of the four quarters, we will drop the value of the first quarter, which is 350, from the total and add the value of the fifth quarter (in other words, first quarter of the next year), and this total will be placed in the middle of the next two values, which are 450 and 400, and so on. These values of the moving totals are shown in column 4 of the next table.

Step 2. The next step is to calculate the quarter moving average. This can be done by dividing the four-quarter moving total, as calculated in Step 1 above, by 4, since there are 4 quarters. The quarters' moving average is recorded in column 5 in the table. The entire table of calculations is shown below:

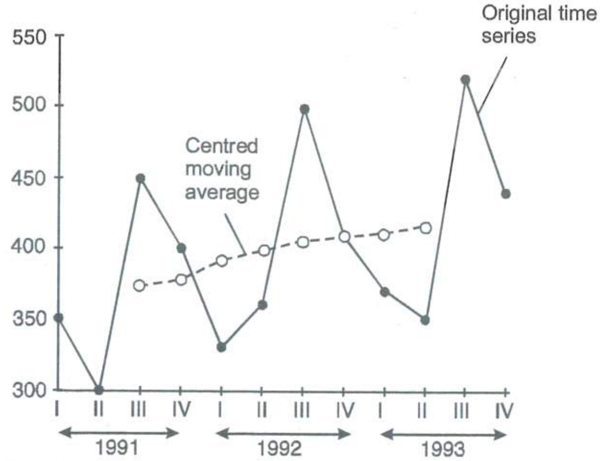
Year	*Quarters*	*Rentals*	*Quarter Moving Total*	*Quarter Moving Average*	*Quarter-centred Moving Average*	*Percentage of Actual to Centred Moving Average*
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1991	I	350				
	II	300				
			1500	375.0		
	III	450			372.50	120.80
			1480	370.0		
	IV	400			377.50	105.96
			1540	385.0		
1992	I	330			391.25	84.35
			1590	397.5		
	II	360			398.75	90.28
			1600	400.0		
	III	500			405.00	123.45
			1640	410.0		
	IV	410			408,75	100.30
			1630	407.5		
1993	I	370			410.00	90.24
			1650	412.5		
	II	350			416.25	84.08
			1680	420.0		
	III	520				
	IV	440				

Step 3. After the moving averages for each consecutive 4 quarters have been taken, then we centre these moving averages. As we see from the above table, the quarterly moving average falls between the quarters. This is because the number of quarters is even which is 4. If we had odd number of time periods, such as 7 days of the week, then the moving average would already be centred and the third step here would not be necessary. Accordingly, we centre our averages in order to associate each average with the corresponding quarter, rather than between the quarters. This is shown in column 6, where the centred moving average is calculated as the average of the two consecutive moving averages.

The moving average (or the centred moving average) aims to eliminate seasonal and irregular fluctuations (S and I) from the original time series, so that this average represents the cyclical and trend components of the series.

NOTES

As the graph shows for this data, the centred moving average has smoothed the peaks and troughs of the original time series.



Step 4. Column 7 in the table contains calculated entries which are percentages of the actual values to the corresponding centred moving average values. The first four quarters' centred moving average of 372.50 in the table has the corresponding actual value of 450, so that the percentage of actual value to centred moving average would be:

$$\frac{\text{Actual value}}{\text{Centred moving average value}} \times 100$$

$$= \frac{450}{372.5} \times 100 \quad = 120.80$$

Step 5. The purpose of this step is to eliminate the remaining cyclical and irregular fluctuations still present in the values in column 7 of the table. This can be done by calculating the 'modified mean' for each quarter. The modified mean for each quarter of the three years time period under consideration, is calculated as follows.

(a) Make a table of values in column 7 of the table given on the percentage of actual to moving average values for each quarter of the three years as shown in the table.

Year	*Quarter I*	*Quarter II*	*Quarter III*	*Quarter IV*
1991	–	–	120.80	105.96
1992	84.35	90.28	123.45	100.30
1993	90.24	84.08	–	–

(b) We take the average of these values for each quarter. It should be noted that if there are many years and quarters taken into consideration instead of 3 years as we have taken, then the highest and lowest values from each quarterly data would be discarded and the average of the remaining data would be considered. By discarding the highest and lowest values from each quarter data, we tend to reduce the extreme cyclical and irregular fluctuations, which are further smoothed when we average the remaining values. Thus, the modified mean can be considered as an index of seasonal component. This modified mean for each quarter data is shown below:

NOTES

$$\text{Quarter I} = \frac{84.35 + 90.24}{2} = 87.295$$

$$\text{Quarter II} = \frac{90.28 + 84.08}{2} = 87.180$$

$$\text{Quarter III} = \frac{120.80 + 123.45}{2} = 122.125$$

$$\text{Quarter IV} = \frac{105.96 + 100.30}{2} = 103.13$$

$$\text{Total} = 399.73$$

The modified means as calculated above are preliminary seasonal indices. These should average 100 per cent or a total of 400 for the 4 quarters. However, our total is 399.73. This can be corrected by the following step.

Step 6. First, we calculate an adjustment factor. This is done by dividing the desired or expected total of 400 by the actual total obtained as 399.73, so that:

$$\text{Adjustment} = \frac{400}{399.73} = 1.0007$$

By multiplying the modified mean for each quarter by the adjustment factor, we get the seasonal index for each quarter as follows:

$$\text{Quarter I} = 87.295 \times 1.0007 = 87.356$$

$$\text{Quarter II} = 87.180 \times 1.0007 = 87.241$$

$$\text{Quarter III} = 122.125 \times 1.0007 = 122.201$$

$$\text{Quarter IV} = 103.13 \times 1.0007 = 103.202$$

$$\text{Total} = 400.000$$

$$\text{Average Seasonal Index} = \frac{400}{4} = 100$$

This average seasonal index is approximated to 100 because of rounding-off errors.

The logical meaning behind this method is based on the fact that the centred moving average part of this process eliminates the influence of secular trend and cyclical fluctuations ($T \times C$). This may be represented by the following expression:

$$\frac{T \times S \times C \times I}{T \times C} = S \times I$$

Where ($T \times S \times C \times I$) is the influence of trend, seasonal variations, cyclic fluctuations and irregular or chance variations.

Thus, the ratio to moving average represents the influence of seasonal and irregular components. However, if these ratios for each quarter over a period of years are averaged, then most random or irregular fluctuations would be eliminated so that,

$$\frac{S \times I}{I} = S$$

and this would give us the value of seasonal influences.

2.2.5 Measuring Irregular Variation

NOTES

Typically, irregular variation is random in nature, unpredictable and occurs over comparatively short periods of time. As a result of its unpredictability, it is generally not measured or explained mathematically. Usually, subjective and logical reasoning explains such variation that occurs. As an example, the current cold weather in Brazil and Columbia is being considered responsible for increase in the price of coffee beans, because cold weather destroys coffee plants. Similarly, the Persian Gulf War, an irregular factor resulted in the increase in airline and ship travel for a number of months because of the movement of personnel and supplies. However, the irregular component can be isolated by eliminating other components from the time series data. The time series data contains $(T \times S \times C \times I)$ components and if we can eliminate $(T \times S \times C)$ elements from the data, then we are left with (I) component. We can follow the previous example to determine the (I) component as follows. The data presented below has already been calculated earlier.

Year	*Quarters*	*Rentals Time Series Values* $(T \times S \times C \times I)$	*Centred Moving Average* $(T \times C)$	$T \times S \times C \times I/(T \times C) = S \times I$
1991	I	350	–	–
	II	300	–	–
	III	450	372.50	1.208
	IV	400	377.50	1.060
1992	I	330	391.25	0.843
	II	360	398.75	0.903
	III	500	405.00	1.235
	IV	410	408.75	1.003
1993	I	370	410.00	0.902
	II	350	416.25	0.841
	III	520	–	–
	IV	440	–	–

The seasonal indices for each quarter have already been calculated as:

Quarter I = 87.356

Quarter II = 87.241

Quarter III = 122.201

Quarter IV = 103.202

Then, the seasonal influence is given by:

Quarter I = 87.356/100 = 0.874

Quarter II = 87.241/100 = 0.872

Quarter III = 122.201/100 = 1.222

Quarter IV = 103.202/100 = 1.032

The given table shows the $(S \times I)$ values, (S) values and the values of (I) calculated by dividing $(S \times I)$ by (S).

NOTES

Year	Quart	(S × I)	(S)	(I)
1991	I	–	–	–
	II	–	–	–
	III	1.208	1.222	0.988
	IV	1.060	1.032	1.027
1992	I	0.843	0.874	0.965
	II	0.903	0.872	1.036
	III	1.235	1.222	1.011
	IV	1.003	1.032	0.972
1993	I	0.902	0.874	1.032
	I	0.841	0.872	0.964
	III	–	–	–
	IV	–	–	–

Seasonal Adjustments

Many times we read about time series values as seasonally adjusted. This is accomplished by dividing the original time series values by their corresponding seasonal indices. These deseasonalized values allow more direct and equitable comparisons of values from different time periods. As for example, in comparing the demands for rental row boats (example that we have evaluated earlier), it would not be equitable to compare the demand of second quarter (spring) with the demand of third quarter (summer), when the demand is traditionally higher. However, these demand values can be compared when we remove the seasonal influence from these time series values.

The seasonally-adjusted values for the demand of row boats in each quarter are based on the values previously calculated.

The following table shows the seasonally-adjusted values:

Year	*Quarters*	*Rentals (T × S × C × I)*	*Seasonal Variation (S)*	*Seasonally-Adjusted Values*	*Rounded-Off Values*
1991	I	350	–	–	–
	II	300	–	–	–
	III	450	1.222	368.25	368
	IV	400	1.032	387.60	388
1992	I	330	0.874	377.57	378
	II	360	0.872	412.80	413
	III	500	1.222	409.16	409
	IV	410	1.032	397.29	397
1993	I	370	0.874	423.34	423
	II	350	0.872	401.38	401
	III	520	–	–	–
	IV	440	–	–	–

The seasonally-adjusted value for each quarter is calculated as:

$$\text{Seasonally-adjusted Value} = \frac{\text{Original value}}{\text{Seasonal index}}$$

These calculations complete the process of separating and identifying the four components of the time series, namely secular trend (T), seasonal variation (S), cyclic variation (C) and irregular variation (I).

NOTES

Check Your Progress

1. What do you mean by the term trend?
2. Give the meaning of cyclical fluctuation.
3. What are the factors that cause seasonal variations?
4. What is irregular random variation?
5. What are the two methods adopted in smoothing techniques?
6. State the methods used to measure seasonal variation.

2.3 PROBABILITY CONCEPTS

Probability (usually represented by the symbol 'P') may be defined as the percentage of times in which a specific outcome would happen if an event was repeated a very large number of times. In other words, the probability of the occurrence of an event is the ratio of the number of times the event occurs (or can occur) to the number of times it and all other events occur (or can occur).

The general meaning of the word 'probability' is likelihood. Where the happening of an event is certain, the probability is said to be unity, i.e., equal to 1 and where there is an absolute impossibility of the happening of an event, the probability is said to be 0. But in real life such cases are rare and the probability generally lies between 0 and 1. Thus, probabilities are always greater than or equal to 0 (i.e., probabilities are never negative) and are equal to or less than 1. This being so, we can say that the weight scale of probability runs from 0 to 1 and in symbolic form it can be stated as:

$$P \leq 1 \text{ but } \geq 0$$

Probability can be expressed either in terms of a fraction or a decimal or a percentage but generally it is expressed in decimals.

2.3.1 The Concept of Sample Space, Sample Points and Events

A *sample space* refers to the complete set of outcomes for the situation as it did or may exist. An element in a set serving as a sample space is called a *sample point*. An *event* is a statement which refers to a particular subset of a sample space for an experiment. The meaning of these three concepts can be easily understood by means of an example. Let us consider an experiment of tossing first one coin and then another. The sample space relevant to it would then consist of all the outcomes of this experiment and can be stated as under:

$$S = [HH, HT, TH, TT]$$

It may be noted that S = [ ] is the symbol used to represent the sample space. This sample space has four outcomes or what we call sample points, viz. *HH*, *HT*, *TH*, and *TT*. One or more of these sample points are called an event. One event may be that both coins fall alike and this can be represented as

$$E_1 = [HH, TT; \text{alike}]$$

The word following the semicolon explains the characteristic of our interest. If E_1 be the event of our interest and E_2 be the subset of all the remaining outcomes, then we have the following equation:

$$S = E_1 + E_2$$

Events

An event is an outcome or a set of outcomes of an activity or a result of a trial. As for example, getting two heads in the trial of tossing three fair coins simultaneously would be an event.

Elementary event

An elementary event, also known as a *simple event*, is a single possible outcome of an experiment. If, for example, we toss a fair coin, then the event of a head coming up is an elementary event. If the symbol for an elementary event is (E), then the probability of the event (E) is written as P[E].

Joint event

A joint event, also known as *compound event,* has two or more elementary events in it. Drawing a *black ace* from a pack of cards, for example, would be a joint event, since it contains two elementary events of black and ace.

Complement of an event

The complement of any event A is the collection of outcomes that are not contained in A. This complement of A is denoted as A′ (A prime). This means that the outcomes contained in A and the outcomes contained in A′ must equal the total sample space. Therefore,

$$P[A] + P[A'] = P[S] = 1$$

or

$$P[A] = 1 - P[A']$$

If for example, a passenger airliner has 300 seats and it is nearly full, but not totally full then event A would be the number of occupied seats and A′ would be the number of unoccupied seats. Suppose, there are 287 seats occupied by passengers and only 13 seats are empty. Typically, the stewardess will count the number of empty seats which are only 13 and report that 287 people are aboard. This is much simpler than counting 287 occupied seats. Accordingly, in such a situation, knowing event A′ is much more efficient than knowing event A.

Mutually exclusive events

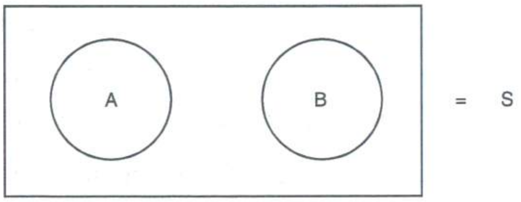
Two events are said to be mutually exclusive, if both events cannot occur at the same time as the outcome of a single experiment. If, for example, we toss a coin, then either event head or event tail would occur, but not both. Hence, these are mutually exclusive events.

NOTES

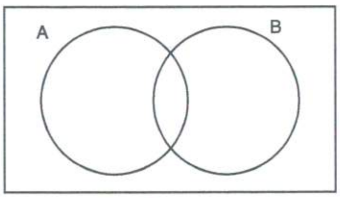
2.3.2 Venn Diagram

A convenient way to visualize the concept of events, their relationships and sample space is done by using Venn diagrams. The sample space is represented by a rectangular region and the events and the relationships among these events are represented by circular regions within the rectangle.

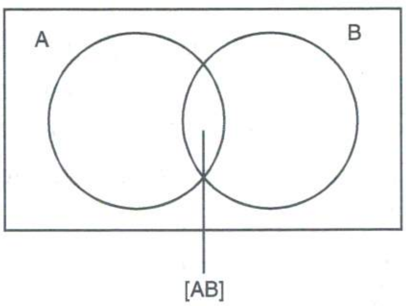
Two mutually exclusive events A and B, for example, are represented in the Venn diagram as follows:



Event $P[A \cup B]$ is represented in the Venn diagram as follows:



Event [AB] is represented as follows:



Union of Three Events

The process of combining two events to form the union can be extended to three events so that $P[A \cup B \cup C]$ would be the union of events A, B, and C. This union can be represented in a Venn diagram.

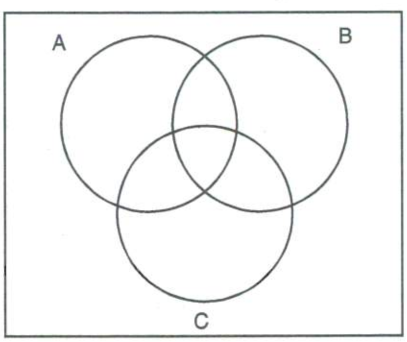


Fig. 2.4 *Union of Three Events*

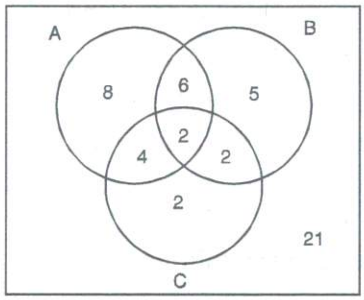
NOTES

***Example 2.4*:** Suppose a sample of 50 students is taken and a survey is made of this sample regarding their reading habits. The survey results are shown as follows:

Event	*Number of Students*	*Magazine they Read*
[A]	20	Time
[B]	15	Newsweek
[C]	10	Filmfare
[AB]	8	Time and Newsweek
[AC]	6	Time and Filmfare
[BC]	4	Newsweek and Filmfare
[ABC]	2	Time and Newsweek and Filmfare

Find out the probability that a student picked up at random from this sample of 50 students who do not read any of these 3 magazines.

***Solution*:** The problem can be solved by a Venn diagram as follows:



Since, there are 21 students who do not read any of the three magazines, the probability that a student picked up at random among this sample of 50 students who do not read any of these three magazines is 21/50.

The problem can also be solved by the formula for probability for union of three events, given as follows:

$$P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[AB] - P[AC] - P[BC] + P[ABC]$$

$$= 20/50 + 15/50 + 10/50 - 8/50 - 6/50 - 4/50 + 2/50$$

$$= 29/50$$

This is the probability that a student picked up at random among the sample of 50 who reads either Time or Newsweek or Filmfare or any combination of the two or all three. Hence, the probability that such a student does not read any of these three magazines is 21/50, which is [1 – 29/50].

2.3.3 Marginal, Conditional and Joint Probabilities

In this section, we will study about marginal, conditional and joint probabilities.

(i) Joint Probability

The joint probability refers to the phenomenon of occurrence of two or more simple events. For example, assume that event (E) is a joint event (or compound event) of drawing a black ace from a pack of cards. There are two simple events involved in the compound event, which are, the card being black and the card being an ace. Hence P [Black ace] or P [E] = 2/52 since there are two black aces in the pack.

NOTES

(ii) Marginal Probability

Marginal probability refers to the unconditional probability $P(A)$ of the event A; i.e., the probability of A, regardless of whether event B did or did not occur. If B can be considered as the event of a random variable X having a given outcome, then the marginal probability of A can be obtained by summing up the joint probabilities over all outcomes for X. For example, if there are two possible outcomes for X with corresponding events B and B', this means that,

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

This is called ***marginalization***. In the above definition, A may precede B or vice versa or they may happen at the same time.

(iii) Conditional Probability

In many situations, a manager may know the outcome of an event that has already occurred and may want to know the chances of a second event occurring based upon the knowledge of the outcome of the earlier event. We are interested in finding out as to how additional information obtained as a result of the knowledge about the outcome of an event affects the probability of the occurrence of the second event. Let us assume, for example, that a new brand of toothpaste is being introduced in the market. Based on the study of competitive markets, the manufacturer has some idea about the chances of its success. Now, he introduces the product in a few selected stores in some selected areas before marketing it nationally. A highly positive response from the test-market area will improve his confidence about the success of his brand nationally. Accordingly, the manufacturer's assessment of high probability of sales for his brand would be conditional upon the positive response from the test-market.

Let there be two events A and B. Then, the probability of event A given the outcome of event B is given by:

$$P[A/B] = \frac{P[AB]}{P[B]}$$

where P[A/B] is interpreted as the probability of event A on the condition that event B has occurred and P [AB] is the joint probability of event A and event B, and P[B] is not equal to zero.

As an example, let us suppose that we roll a die and we know that the number that came up is larger than 4. We want to find out the probability that the outcome is an even number given that it is larger than 4.

Let event A = even

and event B = larger than 4

Then, $$P[\text{even / larger than 4}] = \frac{P[\text{even and larger than 4}]}{P[\text{larger than 4}]}$$

NOTES

or $$P[A/B] = \frac{P[AB]}{P[B]} = \frac{\left(\frac{1}{6}\right)}{\left(\frac{2}{6}\right)} = 1/2$$

But for independent events, P[AB] = P[A]P[B]. Thus, substituting this relationship in the formula for conditional probability, we get:

$$P[A/B] = \frac{P[AB]}{P[B]} = \frac{P[A]P[B]}{P[B]} = P[A]$$

This means that P[A] will remain the same no matter what the outcome of event B is. If, for example, we want to find out the probability of a head on the second toss of a fair coin, given that the outcome of the first toss was a head, this probability would still be 1/2 because the two events are independent events and the outcome of the first toss does not affect the outcome of the second toss.

2.3.4 Addition Theorem of Probability

When two events are mutually exclusive, then the probability that either of the events will occur is the sum of their separate probabilities. If, you roll a single die then the probability that it will come up with a face 5 or face 6, where event A refers to face 5 and event B refers to face 6, both events being mutually exclusive events, is given by

$$P[A \text{ or } B] = P[A] + P[B]$$

or $$P[5 \text{ or } 6] = P[5] + P[6]$$

$$= 1/6 + 1/6$$

$$= 2/6 = 1/3$$

P [A or B] is written as $P[A \cup B]$ and is known as P [A union B].

However, if events A and B are not mutually exclusive, then the probability of occurrence of either event A or event B or both is equal to the probability that event A occurs plus the probability that event B occurs minus the probability that events common to both A and B occur.

Symbolically, it can be written as:

$$P[A \cup B] = P[A] + P[B] - P[A \text{ and } B]$$

P[A and B] can also be written as $P[A \cap B]$, known as P [A intersection B] or simply P[AB].

Events [A and B] consist of all those events which are contained in both A and B simultaneously. As for example, in an experiment of taking cards out of a pack of 52 playing cards, assume that:

Event A = An ace is drawn

Event B = A spade is drawn

Event [AB] = An ace of spade is drawn

Hence, $P[A \cup B] = P[A] + P[B] - P[AB]$

$$= 4/52 + 13/52 - 1/52$$

$= 16/52 = 4/13$

This is because there are 4 aces, 13 cards of spades, including 1 ace of spades out of a total of 52 cards in the pack. The logic behind subtracting P[AB] is that the ace of spades is counted twice – once in event A (4 aces) and once again in event B (13 cards of spade including the ace).

Another example for $P[A \cup B]$, where event A and event B are not mutually exclusive is as follows:

Suppose a survey of 100 persons revealed that 50 persons read *India Today* and 30 persons read *Time* magazine and 10 of these 100 persons read both *India Today* and *Time.* Then,

$$\text{Event } [A] = 50$$
$$\text{Event } [B] = 30$$
$$\text{Event } [AB] = 10$$

Since event [AB] of 10 is included twice, both in event A as well as in event B, event [AB] must be subtracted once in order to determine the event $[A \cup B]$ which means that a person reads *India Today* or *Time* or both. Hence,

$$P[A \cup B] = P[A] + P[B] - P[AB]$$
$$= 50/100 + 30/100 - 10/100$$
$$= 70/100 = 0.7$$

2.3.5 Multiplication Theorem of Probability

Multiplication rule is applied when it is necessary to compute the probability if both events A and B will occur at the same time. The multiplication rule is different if the two events are independent as against the two events being not independent.

If events A and B are independent events, then the probability that they both will occur is the product of their separate probabilities. This is a strict condition so that events A and B are *independent,* and only if,

$$P[AB] = P[A] \times P[B], \text{ or}$$
$$= P[A]P[B]$$

If, we toss a coin twice, then the probability that the first toss results in a head and the second toss results in a tail is given by

$$P[HT] = P[H] \times P[T]$$
$$= 1/2 \times 1/2 = 1/4$$

However, if events A and B are not independent, meaning that the probability of occurrence of an event is dependent or conditional upon the occurrence or non-occurrence of the other event, then the probability that they will both occur is given by

$$P[AB] = P[A] \times P[B/\text{given the outcome of } A]$$

This relationship is written as:

$$P[AB] = P[A] \times P[B/A] = P[A]\,P[B/A]$$

where P[B/A] means the probability of event B on the condition that event A has occurred. As an example, assume that a bowl has 6 black balls and 4 white balls. A ball is drawn at random from the bowl. Then a second ball is drawn without replacement of the first ball back in the bowl. The probability of the second ball being black or white would depend upon the result of the first draw as to whether the first ball was black or white. The probability that both these balls are black is given by

P [two black balls] = P [black on 1st draw] × P [black on 2nd draw/black on 1st draw]

$= 6/10 \times 5/9 = 30/90 = 1/3$

This is so because, first there are 6 black balls out of a total of 10, but if the first ball drawn is black then we are left with 5 black balls out of a total of 9 balls.

2.3.6 Bayes' Theorem and its Business Applications

Reverend Thomas Bayes (1702–1761) introduced his theorem on probability which is concerned with a method for estimating the probability of causes which are responsible for the outcome of an observed effect. Being a religious preacher himself as well as a mathematician, his motivation for the theorem came from his desire to prove the existence of God by looking at the evidence of the world that God created. He was interested in drawing conclusions about the causes by observing the consequences. The theorem contributes to the statistical decision theory in revising prior probabilities of outcomes of events based upon the observation and analysis of additional information.

Bayes' theorem makes use of conditional probability formula where the *condition* can be described in terms of the additional information which would result in the *revised probability* of the outcome of an event.

Suppose that there are 50 students in our statistics class out of which 20 are male students and 30 are female students. Out of the 30 females, 20 are Indian students and 10 are foreign students. Out of the 20 male students, 15 are Indians and 5 are foreigners, so that out of all the 50 students, 35 are Indians and 15 are foreigners. This data can be presented in a tabular form as follows:

	Indian	*Foreigner*	*Total*
Male	15	5	20
Female	20	10	30
Total	35	15	50

Based upon this information, the probability that a student picked up at random will be female is 30/50 or 0.6, since there are 30 females in the total class of 50 students. Now, suppose that we are given additional information that the person picked up at random is Indian, then what is the probability that this person is a female? This additional information will result in revised probability or *posterior probability* in the sense that it is assigned to the outcome of the event after this additional information is made available.

Since we are interested in the revised probability of picking a female student at random provided that we know that the student is Indian, let A_1 be the event *female*, A_2 be the event *male* and B the event *Indian*. Then based upon our knowledge of conditional probability, Bayes' theorem can be stated as follows:

$$P(A_1 / B) = \frac{P(A_1)P(B / A_1)}{P(A_1)P(B / A_1) + P(A_2)(P(B / A_2)}$$

In the example discussed here, there are 2 basic events which are A_1 (female) and A_2 (male). However, if there are n basic events, $A_1, A_2, ...A_n$, then Bayes' theorem can be generalized as,

$$P(A_1 / B) = \frac{P(A_1)P(B / A_1)}{P(A_1)P(B / A_1) + P(A_2)(P(B / A_2) + ... + P(A_n)P(B / A_n)}$$

Solving the case of 2 events we have,

$$P(A_1 / B) = \frac{(30/50)(20/30)}{(30/50)(20/30) + (20/50)(15/20)} = 20/35 = 4/7 = 0.57$$

This example shows that while the *prior probability* of picking up a female student is 0.6, the *posterior probability* becomes 0.57 after the additional information that the student is an American is incorporated in the problem.

Another example of application of Bayes' theorem is as follows:

***Example 2.5*:** A businessman wants to construct a hotel in New Delhi. He generally builds three types of hotels. These are 50 rooms, 100 rooms and 150 rooms in a hotel, depending upon the demand for the rooms, which is a function of the area in which the hotel is located, and the traffic flow. The demand can be categorized as low, medium or high. Depending upon these various demands, the businessman has made some preliminary assessment of his net profits and possible losses (in thousands of dollars) for these various types of hotels. These pay-offs are shown in the following table.

States of Nature

Demand for Rooms

		Low (A_1)	Medium (A_2)	High (A_3)	
		0.2	0.5	0.3	Demand Probability
Number of Rooms	$R_1 = (50)$	25	35	50	
	$R_2 = (100)$	–10	40	70	
	$R_3 = (150)$	–30	20	100	

***Solution*:** The businessman has also assigned 'prior probabilities' to the demand structure or rooms. These probabilities reflect the initial judgement of the businessman based upon his intuition and his degree of belief regarding the outcomes

NOTES

of the states of nature.

Demand for rooms	*Probability of Demand*
Low (A_1)	0.2
Medium (A_2)	0.5
High (A_3)	0.3

Based upon these values, the expected pay-offs for various rooms can be computed as follows:

$$EV(50) = (25 \times 0.2) + (35 \times 0.5) + (50 \times 0.3) = 37.50$$

$$EV(100) = (-10 \times 0.2) + (40 \times 0.5) + (70 \times 0.3) = 39.00$$

$$EV(150) = (-30 \times 0.2) + (20 \times 0.5) + (100 \times 0.3) = 34.00$$

This gives us the maximum pay-off of $39,000 for building a hotel with 100 rooms.

Now, the hotelier must decide whether to gather additional information regarding the states of nature, so that these states can be predicted more accurately than the preliminary assessment. The basis of such a decision would be the cost of obtaining additional information. If this cost is less than the increase in maximum expected profit, then such additional information is justified.

Suppose that the businessman asks a consultant to study the market and predict the states of nature more accurately. This study is going to cost the businessman $10,000. This cost would be justified if the maximum expected profit with the new states of nature is at least $10,000 more than the expected pay-off with the prior probabilities. The consultant made some studies and came up with the estimates of low demand (X_1), medium demand (X_2), and high demand (X_3) with a degree of reliability in these estimates. This degree of reliability is expressed as conditional probability which is the probability that the consultant's estimate of low demand will be correct and the demand will be actually low. Similarly, there will be a conditional probability of the consultant's estimate of medium demand, when the demand is actually low, and so on. These conditional probabilities are expressed in Table 2.1.

***Table 2.1** Conditional Probabilities*

		X_1	X_2	X_3
States of	(A_1)	0.5	0.3	0.2
Nature	(A_2)	0.2	0.6	0.2
(Demand)	(A_3)	51	0.3	0.6

The values in the preceding table are conditional probabilities and are interpreted as follows:

The value 0.5 is the probability that the consultant's prediction will be for low demand (X_1) when the demand is actually low. Similarly, the probability is 0.3 that the consultant's estimate will be for medium demand (X_2) when in fact the demand is low, and so on. In other words, $P(X_1/A_1) = 0.5$ and $P(X_2/A_1) = 0.3$. Similarly, $P(X_1/A_2) = 0.2$ and $P(X_2/A_2) = 0.6$, and so on.

NOTES

Our objective is to obtain posteriors which are computed by taking the additional information into consideration. One way to reach this objective is to first compute the joint probability which is the product of prior probability and conditional probability for each state of nature. Joint probabilities, as computed, is given as –

State of Nature	*Prior Probability*	*Joint Probabilities* $P(A_1X_1)$	$P(A_1X_2)$	$P(A_1X_3)$
A_1	0.2	$0.2 \times 0.5 = 0.1$	$0.2 \times 0.3 = 0.06$	$0.2 \times 0.2 = 0.04$
A_2	0.5	$0.5 \times 0.2 = 0.1$	$0.5 \times 0.6 = 0.3$	$0.5 \times 0.2 = 0.1$
A_3	0.3	$0.3 \times 0.1 = 0.03$	$0.3 \times 0.3 = 0.09$	$0.3 \times 0.6 = 0.18$
Total marginal probabilities		$= 0.23$	$= 0.45$	$= 0.32$

Now, the posterior probabilities for each state of nature A_i are calculated as follows:

$$P(A_i / X_j) = \frac{\text{Joint probability of } A_i \text{ and } X_j}{\text{Marginal probability of } X_j}$$

By using this formula, the joint probabilities are converted into posterior probabilities and the computed table for these posterior probabilities is given below.

States of Nature	*Posterior Probabilities* $P(A_1/X_1)$	$P(A_1/X_2)$	$P(A_1/X_3)$
A_1	$0.1/.023 = 0.435$	$0.06/0.45 = 0.133$	$0.04/0.32 = 0.125$
A_2	$0.1/.023 = 0.435$	$0.30/0.45 = 0.667$	$0.1/0.32 = 0.312$
A_3	$0.03/.023 = 0.130$	$0.09/0.45 = 0.200$	$0.18/0.32 = 0.563$
Total	$= 1.0$	$= 1.0$	$= 1.0$

Now, we have to compute the expected pay-offs for each course of action with the new posterior probabilities assigned to each state of nature. The net profits for each course of action for a given state of nature is the same as before and is restated as follows. These net profits are expressed in thousands of dollars.

		Low (A_1)	*Medium* (A_2)	*High* (A_3)
Number of Rooms	(R_1)	25	35	50
	(R_2)	–10	40	70
	(R_3)	–30	20	100

Let O_{ij} be the monetary outcome of course of action (i) when (j) is the corresponding state of nature, so that in the above case O_{i1} will be the outcome of course of action R_1 and state of nature A_1, which in our case is \$25,000. Similarly, O_{i2} will be the outcome of action R_2 and state of nature A_2, which in our case is \$10,000, and so on. The expected value EV (in thousands of dollars) is calculated on the basis of the actual state of nature that prevails as well as the estimate of the state of nature as provided by the consultant. These expected values are calculated as follows:

Course of action $= R_i$

Estimate of consultant $= X_i$

Actual state of nature $= A_i$

where $i = 1, 2, 3$

Then,

NOTES

(A) Course of action $= R_1$ = Build 50-rooms-hotel

$$EV\left(\frac{R_1}{X_1}\right) = \Sigma P\left(\frac{A_i}{X_1}\right) O_{i1}$$
$$= 0.435(25) + 0.435(-10) + 0.130(-30)$$
$$= 10.875 - 4.35 - 3.9 = 2.625$$

$$EV\left(\frac{R_1}{X_2}\right) = \Sigma P\left(\frac{A_i}{X_2}\right) O_{i1}$$
$$= 0.133(25) + 0.667(-10) + 0.200(-30)$$
$$= 3.325 - 6.67 - 6.0 = -9.345$$

$$EV\left(\frac{R_1}{X_3}\right) = \Sigma P\left(\frac{A_i}{X_3}\right) O_{i1}$$
$$= 0.125(25) + 0.312(-10) + 0.563(-30)$$
$$= 3.125 - 3.12 - 16.89$$
$$= -16.885$$

(B) Course of action $= R_2$ = Build 100-rooms-hotel

$$EV\left(\frac{R_2}{X_1}\right) = \Sigma P\left(\frac{A_i}{X_1}\right) O_{i2}$$
$$= 0.435(35) + 0.435(40) + 0.130(20)$$
$$= 15.225 + 17.4 + 2.6 = 35.225$$

$$EV\left(\frac{R_2}{X_2}\right) = \Sigma P\left(\frac{A_i}{X_1}\right) O_{i2}$$
$$= 0.133(35) + 0.667(40) + 0.200(20)$$
$$= 4.655 + 26.68 + 4.0 = 35.335$$

$$EV\left(\frac{R_2}{X_3}\right) = \Sigma P\left(\frac{A_i}{X_3}\right) O_{i2}$$
$$= 0.125(35) + 0.312(40) + 0.563(20)$$
$$= 4.375 + 12.48 + 11.26 = 28.115$$

(C) Course of action $= R_3$ = Build 150-rooms-hotel

$$EV\left(\frac{R_3}{X_1}\right) = \Sigma P\left(\frac{A_i}{X_1}\right) O_{i3}$$
$$= 0.435(50) + 0.435(70) + 0.130(100)$$
$$= 21.75 + 30.45 + 13 = 65.2$$

$$EV\left(\frac{R_3}{X_2}\right) = \Sigma P\left(\frac{A_i}{X_2}\right)O_{i3}$$
$$= 0.133(50) + 0.667\,(70) + 0.200\,(100)$$
$$= 6.65 + 46.69 + 20 = 73.34$$

NOTES

$$EV\left(\frac{R_3}{X_3}\right) = \Sigma P\left(\frac{A_i}{X_3}\right)O_{i3}$$
$$= 0.125(50) + 0.312(70) + 0.563(100)$$
$$= 6.25 + 21.84 + 56.3 = 84.39$$

The calculated expected values in thousands of dollars are presented in a tabular form.

Table 2.2 *Expected Posterior Pay-offs*

Outcome	$EV(R_1/X_i)$	$EV(R_2/X_i)$	$EV(R_3/X_i)$
X_1	2.625	35.225	65.2
X_2	–9.345	35.335	73.34
X_3	–16.885	28.115	84.39

Table 2.2 can now be analysed in the following manner.

If the outcome is X_1, it is desirable to build-150-rooms-hotel, since the expected pay-off for this course of action is maximum of $65,200. Similarly, if the outcome is X_2, the course of action should again be R_3 since the maximum pay-off is $73,34. Finally, if the outcome is X_3, the maximum pay-off is $84,390 for course of action R_3.

Accordingly, given these conditions and the pay-off, it would be advisable to build a hotel which has 150 rooms.

Check Your Progress

7. Define Probability.
8. What is a mutually exclusive event?
9. Mention the concept of independent events.
10. What is addition rule?
11. When is the law of multiplication applied?
12. What is Bayes' theorem?

2.4 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The term trend means the general long-term movement in the time series value of the variable (*Y*) over a fairly long period of time. Here, '*Y*' stands for factors like sales, population, crime rate, etc., evaluated for the future analysis and forecasting.

NOTES

2. Regular swings or patterns that repeat over a long period of time are known as cyclical fluctuations. These are usually unpredictable in relation to the time of occurrence, duration and amplitude.
3. Factors like changes in climate and weather, and customs and traditions cause seasonal variations.
4. Variations that are accidental, random or occur due to chance factors, are known as irregular random variations.
5. The two methods adopted in smoothing techniques are:
   (a) Moving averages
   (b) Exponential smoothing
6. The methods used to measure seasonal variation are:
   (a) Simple average method
   (b) Ratio to moving average method
7. Probability (usually represented by the symbol 'P') may be defined as the percentage of times in which a specific outcome would happen if an event was repeated a very large number of times. The general meaning of the word 'probability' is likelihood.
8. Two events are said to be mutually exclusive, if both events cannot occur at the same time as the outcome of a single experiment.
9. Two events A and B are said to be independent events, if the occurrence of one event is not at all influenced by the occurrence of the other.
10. The addition rule states that when two events are mutually exclusive, then the probability that either of the events will occur is the sum of their separate probabilities.
11. Multiplication rule is applied when it is necessary to compute the probability in case two events occur at the same time.
12. Bayes' theorem on probability is concerned with a method for estimating the probability of causes which are responsible for the outcome of an observed effect. The theorem contributes to the statistical decision theory in revising prior probabilities of outcomes of events based upon the observation and analysis of additional information.

2.5 SUMMARY

- The time series analysis method is quite accurate where the future is expected to be similar to the past. The underlying assumption in time series is that the same factors will continue to influence the future patterns of economic activity in a similar manner as in the past.
- Secular trend is the general long-term movement in the time series value of the variable (Y) over a fairly long period of time. The variable (Y) is the factor that we consider while making evaluation for the future. It could be sales, population, crime rate, etc.

NOTES

- If a trend can be determined and the rate of change can be ascertained, then tentative estimates on the same series of values into the future can be made.
- The cyclical fluctuations refer to regular swings or patterns that repeat over a long period of time. The movements are considered cyclical only if they occur after time intervals of more than one year. These are the changes that take place as a result of economic booms or depressions.
- Smoothing techniques improve the forecasts of future trends provided that the time series are fairly stable with no significant trend, cyclical or seasonal effect and the objective is to *smooth out* the irregular component of the time series through the averaging process.
- The concept of the moving averages is based on the idea that any large irregular component of time series at any point in time will have a less significant impact on the trend, if the observation at that point in time is averaged with such values immediately before and after the observation under consideration.
- In the moving average method, each observation in the moving average calculation receives the same weight.
- When a time series shows an upward or downward long-term linear trend, then regression analysis can be used to estimate this trend and project the trends into forecasting the future values of the variables involved.
- Cyclic variation is a pattern that repeats over time periods longer than one year. These variations are generally unpredictable in relation to the time of occurrence, duration as well as amplitude.
- The measure used to identify cyclical variation is the percentage of trend and the procedure used is known as the residual trend.
- Seasonal variation has been defined as predictable and repetitive movement around the trend line in a period of one year or less. For the measurement of seasonal variation, the time interval involved may be in terms of days, weeks, months or quarters.
- Simple average method is the simplest method of isolating seasonal fluctuations in time series. It is based on the assumption that the series contain only the seasonal and irregular fluctuations.
- Probability (usually represented by the symbol 'P') may be defined as the percentage of times in which a specific outcome would happen if an event was repeated a very large number of times.
- The general meaning of the word 'probability' is likelihood. Where the happening of an event is certain, the probability is said to be unity, i.e., equal to 1 and where there is an absolute impossibility of the happening of an event, the probability is said to be 0.
- Probability can be expressed either in terms of a fraction or a decimal or a percentage but generally it is expressed in decimals.

- A sample space refers to the complete set of outcomes for the situation as it did or may exist. An element in a set serving as a sample space is called a sample point. An event is a statement which refers to a particular subset of a sample space for an experiment.
- Two events are said to be mutually exclusive, if both events cannot occur at the same time as the outcome of a single experiment.
- A convenient way to visualize the concept of events, their relationships and sample space is done by using Venn diagrams. The sample space is represented by a rectangular region and the events and the relationships among these events are represented by circular regions within the rectangle.
- The joint probability refers to the phenomenon of occurrence of two or more simple events.
- Multiplication rule is applied when it is necessary to compute the probability if both events A and B will occur at the same time. The multiplication rule is different if the two events are independent as against the two events being not independent.

2.6 KEY TERMS

- **Cyclic fluctuation:** It refers to the regular swings or patterns that repeat over a long period of time, i.e., periods longer than one year.
- **Seasonal variation:** It refers to the patterns of change that repeat over a period of one year or less.
- **Irregular variation:** Irregular variations are unpredictable and can be accidental, random or simply due to chance factor.
- **Smoothing techniques:** It is used to smooth out the irregular components of the time series and improve the forecasts of future trends using the averaging process.
- **Seasonal adjustment:** It is calculated by dividing the original time series values by their corresponding seasonal indices.
- **Classical theory of probability:** It is the theory of probability based on the number of favourable outcomes and the number of total outcomes.
- **Event:** It is an outcome or a set of outcomes of an activity or the result of a trial.
- **Elementary event:** It is the single possible outcome of an experiment. It is also known as a simple event.
- **Joint event:** Joint event has two or more elementary events in it and is also known as compound event.
- **Sample space:** It is the collection of all possible events or outcomes of an experiment.
- **Addition rule:** It states that when two events are mutually exclusive, then the probability that either of the events will occur is the sum of their separate probabilities.

- **Multiplication rule:** It is applied when it is necessary to compute the probability, if both events *A* and *B* occur at the same time. Different rules are applied for different conditions.

NOTES

2.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What do you mean by trend analysis?
2. Differentiate between secular trend and cyclic fluctuations.
3. How is irregular variation caused?
4. Define seasonal variation.
5. How is the cyclical effect measured?
6. What is the simple average method of isolating seasonal fluctuations in time series?
7. What are the ways to measure irregular variation?
8. How the seasonal adjustments made?
9. What do you mean by sample space, sample points and events?
10. What is a mutually exclusive event?
11. What is Venn diagram?
12. What do you mean by marginal, conditional and joint probabilities?
13. What is the addition theorem of probability?
14. State the concept of multiplication rule.
15. What is Bayes' theorem? What is its importance in statistical calculations?

Long-Answer Questions

1. The following data shows the number of Lincoln Continental cars sold by a dealer in Queens during the 12 months of 1994.

Months	*Number Sold*
January	52
February	48
March	57
April	60
May	55
June	62
July	54
August	65
September	70
October	80
November	90
December	75

NOTES

(a) Calculate the 3-month moving average for this data.

(b) Calculate the 5-month moving average for this data.

(c) Which one of these two moving averages is a better smoothing technique and why?

2. An economist has calculated the variable rate of return on money market funds for the last 12 months as follows:

Months	*Rate of Return (%)*
January	6.2
February	5.8
March	6.5
April	6.4
May	5.9
June	5.9
July	6.0
August	6.8
September	6.5
October	6.1
November	6.0
December	6.0

(a) Using a 3-month moving average, forecast the rate of return for next January.

(b) Using exponential smoothing method and setting = 0.8, forecast the rate of return for next January.

3. An institution dealing with pension funds is interested in buying a large block of stock of Azumi Business Enterprises (ABE). The president of the institution has noted down the dividends paid out on common stock shares for the last 10 years. This data is presented as follows:

Years	*Dividend ($)*
1985	3.20
1986	3.00
1987	2.80
1988	3.00
1989	2.50
1990	2.10
1991	1.60
1992	2.00
1993	1.10
1994	1.00

(a) Plot the data.

(b) Estimate the dividend expected in 1995.

(c) Calculate the points on the trend line for the years 1987 and 1991 and plot the trend line.

NOTES

4. Rinkoo Camera Corporation has ten camera stores scattered in the five areas of New York city. The president of the company wants to find out if there is any connection between the sales price and the sales volume of Nikon F-1 camera in the various retail stores. He assigns different prices of the same camera for the different stores and collects data for a 30-day period. The data is presented as follows. The sales volume is in number of units and the price is in dollars.

Stores	*Price*	*Volume*
1	550	420
2	600	400
3	625	300
4	575	400
5	600	340
6	500	440
7	450	500
8	480	460
9	550	400
10	650	310

(a) Plot the data.

(b) What effect would you expect on sales if the price of the camera in store number 7 is increased to $530?

(c) Calculate the points on the trend line for stores 4 and 7 and plot the trend line.

5. The following table gives a times series of monthly sales of Luxury Automobiles for a big car dealer for each month of the years 1994 and 1995.

Months	*1994*	*1995*
January	720	780
February	800	840
March	1080	1110
April	1000	980
May	1020	1050
June	1105	905
July	900	880
August	930	910
September	830	780
October	1030	920
November	880	770
December	800	730

Derive the seasonal index using the ratio to moving average.

6. The following data represent the index of total industrial production (Y) for each quarter of the last four years.

NOTES

Years	*Quarters*	*(Y)*
1st year	1	103.1
	2	107.2
	3	109.0
	4	102.1
2nd year	1	105.9
	2	109.7
	3	112.1
	4	106.0
3rd year	1	110.0
	2	112.6
	3	112.8
	4	104.3
4th year	1	107.0
	2	105.2
	3	104.8
	4	99.6

Calculate the quarter moving averages, quarter-centred moving averages and percentage of actual to centred moving averages.

7. The sales of pianos at Kiran Musical Enterprise for the last ten months have been 1004, 99, 110, 120, 105, 112, 109, 117, 120 and 125 respectively. The owner of the store is interested in using exponential smoothing to aid in analysing these data. The owner is interested in a greater degree of smoothing and hence he has set the value of the smoothing factor = 0.3. Calculate the expected sale for next month.

8. The following data represent the quarterly earnings per share of a software company for the last four years.

	Quarters			
Years	1	2	3	4
1st year	0.27	0.35	0.43	1.25
2nd year	0.40	0.55	0.45	1.35
3rd year	0.52	0.70	0.53	1.55
4th year	0.60	0.80	0.64	1.85

Analyse the quarterly time series to determine the effects of the trend, cyclic, seasonal and irregular components.

9. Consider the time series of quarterly sales (in thousands of dollars) for a local departmental store for the last 3 years as shown in the following table.

Years	*Quarters*	*Sales*
1993	1	530
	2	520
	3	690
	4	570
1994	1	550
	2	600
	3	840
	4	640
1995	1	650
	2	620
	3	990
	4	800

NOTES

The seasonal indices for each quarter are shown in the following table:

Quarters	Index
1	0.90
2	0.88
3	1.25
4	0.95

Find the seasonally-adjusted sales corresponding to each sales value.

10. The following data represent the values of percentage of actual data to centred moving average for each quarter of the last 6 years for sales of 52 inch screen Sony television of a big appliance store.

	Quarters			
Years	*Fall*	*Winter*	*Spring*	*Summer*
1990	–	–	90	30
1991	150	120	98	35
1992	160	115	95	30
1993	152	108	100	28
1994	145	115	107	32
1995	152	120	–	–

Determine the seasonal index for each quarter.

11. A restaurant manager has recorded the daily number of customers for the last 4 weeks. He wants to improve customer service and change employee scheduling as far as necessary, based on the expected number of daily customers in the future. The following data represent the daily number of customers as recorded by the manager for the last 4 weeks.

Weeks	*Mon*	*Tues*	*Wed*	*Thurs*	*Fri*	*Sat*	*Sun*
1	440	400	480	510	650	800	710
2	510	430	500	520	740	850	800
3	490	480	410	630	720	810	690
4	500	500	470	540	780	900	850

Determine the daily seasonal indices using the 7-day moving average.

12. The Pacific Amusement Park, located in Silicon Valley has provided the following data on the number of visitors (in thousands of admissions) during the Park's Open Seasons of Spring, Summer and Fall.

Years	*Spring*	*Summer*	*Fall*
1991	280	610	220
1992	300	725	180
1993	140	600	200
1994	200	580	180

Calculate the seasonal indices for the above given data.

13. A real estate agency has been in business for the last 4 years and specializes in sales of two-family houses. The sales in the last 4 years have grown from 20 houses in the first year to 105 houses last year. The owner of the agency would like to develop a forecast for sale of houses in the coming year. The quarterly sales data for the last 4 years are shown below:

NOTES

Years	Quarter(1)	Quarter(2)	Quarter(3)	Quarter (4)
1	8	6	2	4
2	10	8	8	12
3	18	12	15	25
4	25	20	28	32

(a) Using moving average method, find the values of combined trend and cyclical component.

(b) Find the values of combined seasonal and irregular component.

(c) Compute the seasonal indices for the 4 quarters.

(d) Deseasonalize the data and use the deseasonalized time series to identify the trend.

(e) Find the value of irregular component.

14. A family plans to have two children. What is the probability that both children will be boys? (List all the possibilities and then select the one which would be two boys.)

15. A card is selected at random from an ordinary well-shuffled pack of 52 cards. What is the probability of getting:

    (a) A king (b) A spade

    (c) A king or an ace (d) A picture card

16. A wheel of fortune has numbers 1 to 40 painted on it, each number being at equal distance from the other so that when the wheel is rotated, there is the same chance that the pointer will point at any of these numbers. Tickets have been issued to contestants numbering 1 to 40. The number at which the wheel stops after being rotated would be the winning number. What is the probability that,

    (a) Ticket number 29 wins.

    (b) One person who bought 5 tickets numbered 18 to 22 (inclusive), wins the prize.

17. The Dean of the School of Business has two secretaries, Mary and Jane. The probability that Mary will be absent on any given day is 0.08. The probability that Jane will be absent on any given day is 0.06. The probability that both the secretaries will be absent on any given day is 0.02. Find the probability that either one of them will be absent on any given day.

18. Two fair dice are rolled. What is the probability of getting:

    (a) A sum of 10 or more

    (b) A pair of which atleast one number is 3

    (c) A sum of 8, 9, or 10

    (d) One number less than 4

19. An urn contains 12 white balls and 8 red balls. Two balls are to be selected in succession, at random and without replacement. What is the probability that

(a) Both balls are white.

(b) The first ball is white and the second ball is red.

(c) One white ball and one red ball are selected.

Would the probabilities change if the first ball after being identified is put back in the urn before the second ball is selected?

NOTES

20. 200 students from the college were surveyed to find out if they were taking any of the Management, Marketing or Finance courses. It was found that 80 of them were taking Management courses, 70 of them were taking Marketing courses and 50 of them were taking Finance courses. It was also found that 30 of them were taking Management and Marketing courses, 30 of them were taking Management and Finance courses and 25 of them were taking Marketing and Finance courses. It was further determined that 20 of these students were taking courses in all the three areas. What is the probability that a particular student is not taking any course in any of these areas?

21. A family plans to have three children. List all the possible combinations and find the probability that all the three children will be boys.

22. A movie house is filled with 700 people and 60% of these people are females. 70% of these people are seated in the no smoking area including 300 females. What is the probability that a person picked up at random in the movie house is:

    (a) A male.

    (b) A female smoker.

    (c) A male or a non-smoker.

    (d) A smoker if we knew that the person is a male.

    (e) Are the events sex and smoking statistically independent?

23. A fair dice is rolled once. What is the probability of getting:

    (a) An odd number

    (b) A number greater than 3

24. In a computer course, the probability that a student will get an A is 0.09. The probability that he will get a B grade is 0.15 and the probability that he will get a C grade is 0.45. What is the probability that the student will get either a D or an F grade?

25. In a statistics class, the probability that a student picked up at random comes from a two parent family is 0.65, and the probability that he will fail the exam is 0.20. What is the probability that such a randomly selected student will be a low achiever given that he comes from a two parent family?

26. The following is a breakdown of faculty members in various ranks at the college.

NOTES

Rank	*Number of Males*	*Number of Females*
Professor	20	12
Assoc. Professor	18	20
Asst. Professor	25	30

What is the probability that a faculty member selected at random is:

(a) A female.

(b) A female professor.

(c) A female given that the person is a professor.

(d) A female or a professor.

(e) A professor or an assistant professor.

(f) Are the events of being a male and being an associate professor statistically independent events?

27. A car dealer in a suburban community is interested to make a survey of the number of cars the families in the community owned. He selected 333 families and recorded the following results.

Number of Cars	*Number of Families*
0	20
1	44
2	170
3	63
4	36

To promote his dealership, the dealer selects a family at random by the lottery method to award two tickets to Puerto Rico. What is the probability that the family selected:

(a) Owns no car.

(b) Owns 2 cars.

(c) Owns 2 cars or more.

(d) Owns only one car given that the families with no cars are excluded from the process.

28. A part-time student is taking two courses, namely Statistics and Finance. The probability that the student will pass the Statistics course is 0.60 and the probability of passing the Finance course is 0.70. The probability that the student will pass both courses is 0.50. Find the probability that the student:

(a) Will pass at least one course

(b) Will pass either or both courses

(c) Will fail both courses

29. In how many ways can a person choose 4 books from a list of 8 best-sellers?

30. There are five finalists in a beauty contest. Three of these finalists are to be selected as winners including the winner of the contest as well as a first

runner-up and a second runner-up. In how many different ways can such a combination be obtained?

31. Out of 20 students in a Statistics class, 3 students fail in the course. If 4 students from the class are picked up at random, what is the probability that one of the failing students will be among them.

NOTES

32. The Psychology class has decided to organize a Christmas party. The class has only 18 students including 12 women. The professor has decided to pick a group of 4 students at random and assign this group the responsibility of making all the arrangements. What is the probability that this group consists of,
    (a) All women
    (b) 2 women and 2 men

33. The New York Pick Five lottery drawing draws five numbers at random out of 39 numbers labelled 1 to 39. How many different outcomes are possible?

34. A company has 18 senior executives. Six of these executives are women including four blacks and two Indians. Six of these executives are to be selected at random for a Christmas cruise. What is the probability that the selection will include:

    (a) All the black and Indian women (b) At least one Indian woman

    (c) Not more than two women (d) Half men and half women

35. An independent insurance company has 27 employees. 15 of them sell life insurance, 7 of them sell automobile insurance and 4 of them sell both life and auto insurance. The others do the administrative work. If one of the employees is selected at random, what is the probability that such a person does administrative work?

36. The job placement office at City University keeps a record of the graduating students who apply for jobs through this office. The record shows that 70 per cent of candidates are graduates and 30 per cent are undergraduates. The record also indicates that a graduate applicant has a 65 per cent chance of getting a job while the chance of an undergraduate being placed on a job is 35 per cent.
    (a) What is the probability that a student randomly coming to the office to apply for a job will get the job?
    (b) A student comes to the office to happily announce that she got the job. What is the probability that she is a graduate student?

37. The Department of Transportation in the city was asked to study the records of all employees who got their training in the city technical institute. It was found that 20 per cent of all such graduates were women and 15 per cent belong to minority groups. Only 10 per cent of the minority graduates were women. Find the probability that a technically trained person, selected at random is:

(a) A member of the minority group
(b) A female member of a non-minority group
(c) A male, given that the member belongs to a minority group
(d) A female or a member of the non-minority group

38. The probability that a management trainee will remain with the company after the training programme is completed is 0.70. The records indicate that 60 per cent of all managers earn over $60,000 per year. The probability that an employee is a management trainee or who earns more than $60,000 per year is 0.80. What is the probability that an employee earns more than $60,000 per year, given that he is a management trainee who stayed with the company after completing the training programme?

39. A meteorologist has forecast the probability of rain on Monday, Tuesday and Wednesday as follows:

Day	*Probability of Rain*
Monday	0.60
Tuesday	0.50
Wednesday	0.30

Assuming that the weather from day to day is independent, what is the probability that it will rain at least once in these three days?

40. An investor buys 100 shares each of the three stocks A, B and C. Based on past statistical analysis, the investor has calculated probabilities of the values of these stocks to increase in one week time period as 0.80, 0.70 and 0.60, respectively. Assuming that the movements of these stocks are independent events, what is the probability that:
    (a) Exactly two of the three stocks will increase in value in the given week.
    (b) At least two stocks will increase in value.
    (c) All three stocks will increase in value.
    (d) No more than one stock will increase in value.

2.8 FURTHER READING

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House.

Gupta, S. C. 2006. *Fundamentals of Statistics*. New Delhi: Himalaya Publishing House.

Gupta, S. P., 2005. *Statistical Methods*. New Delhi: Sultan Chand and Sons.

Hooda, R. P. 2002. *Statistics for Business and Economics*. New Delhi: Macmillan India.

Kothari, C. R., 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Monga, G. S. 2000. *Mathematics and Statistics for Economics*. New Delhi: Vikas Publishing House

Gupta, S.P. 2006. *Statistical Methods*. New Delhi: S. Chand & Co. Ltd.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Gupta, S.C. and V.K. Kapoor. *Fundamentals of Mathematical Statistics*. New Delhi: Sultan Chand & Sons.

Sharma, A. 2009. *Operations Research.* New Delhi: Himalaya Publishing House.

Singh, J. K. 2009. *Business Mathematics*. New Delhi: Himalaya Publishing House.

Kalavathy, S. 2010. *Operations Research with C Programmes,* 3rd edition. New Delhi: Vikas Publishing House Pvt. Ltd.

Kothari, C. R. 1992. *An Introduction to Operational Research.* New Delhi: Vikas Publishing House Pvt. Ltd.

NOTES

UNIT 3 INFERENTIAL DECISION-MAKING AND THE DECISION-MAKING PROCESS

NOTES

Structure

3.0 INTRODUCTION

In this unit, you will learn about the decision-making process. A decision taken by a manager has far-reaching effect on a business. Decisions are of two types; tactical and strategic. Tactical decisions affect the business in the short run whereas strategic decisions have a far-reaching effect. To take a right decision, managers resort to statistical methods to analyse factors that affect the business as a whole. You will know about decision models, deterministic and probabilistic. A lot of foresight is required to take decisions with probabilistic models and for that some rule or criteria have to be followed. A decision-making process adopts EMV (Expected Monetary Value) and EOL (Expected Opportunity Loss) criteria for uncertain situations which involve risk.

You will learn about marginal analysis, which is used as an alternative to EMV or EOL, since these two methods involve lot of computations and are tedious processes. Marginal analysis starts by considering whether an additional unit bought will either be sold or not. This is based on probability theory. You will learn about the competitive decision model which is related to situations are uncertain. There are several rules to decide which are maximin decision rule, maximax decision rule, Salvage decision rule, Hurwicz decision rule and Laplace decision rule.

Finally, you will learn about the decision tree approach. A decision tree is a decision flow diagram. In a decision tree approach, the expectation principles are generally used. Here we choose between alternatives that minimize the expected profit and minimize expected costs.

3.1 OBJECTIVES

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

- Describe the quantitative approach to management decision-making
- Understand the decision-making process
- Describe decision-making under different situations
- Explain the method of preparing payoff tables
- Discuss the deterministic and probabilistic models of decision-making
- Explain marginal analysis
- Understand how decision trees are used to make decisions

3.2 QUANTITATIVE APPROACH TO MANAGEMENT DECISION-MAKING

Decision-making is an everyday process in life. It is the major role of a manager too. The decision taken by a manager has far reaching effect on the business. Right decisions will have salutary effect and the wrong one may prove to be disastrous.

3.2.1 Decision Theory

Decisions may be classified into two categories, tactical and strategic. Tactical decisions are those which affect the business in the short run. Strategic decisions are those which have far reaching effect on the course of business.

These days, in every organization whether large or small, the person at the top management has to take some decision, knowing that certain events beyond his control may occur to make him regret the decision. He is uncertain as to whether or not these unfortunate events will happen. In such situations, the best possible decision can be made by the use of statistical methods. The methods try to minimize the degree to which the person is likely to regret the decision he makes for a particular problem.

Decision-making constitutes one of the highest forms of human activity. Statistics provides tools for making wise decisions in the face of uncertainity. This has led to the development of statistical methods. The methods try to minimize the degree to which the person is likely to regret the decision he makes for a particular problem.

The problem under study may be represented by a model in terms of the following elements:

(*i*) ***The decision-maker.*** The decision-maker is charged with the responsibility of making the decision. That is he has to select one from a set of possible courses of action.

(*ii*) ***The acts.*** The acts are the alternative courses of action of strategies that are available to the decision-maker. The decision involves a selection among two or more alternative courses of action. The problem is to choose the best of these alternatives to achieve an objective.

(*iii*) ***Event.*** Events are the occurrences which affect the achievement of the objectives. They are also called states of nature or outcomes. The events constitute a mutually exclusive and exhaustive set of outcomes, which describe the possible behaviour of the environment in which the decision is made. The decision-maker has no control over which event will take place and can only attach a subjective probability of occurrence of each.

(*iv*) ***Payoff table.*** A payoff table represents the economics of a problem, i.e., revenue and costs associated with any action with a particular outcome. It is an ordered statement of profit or costs resulting under the given situation. The payoff can be interpreted as the outcome in quantitative form if the decision-maker adopts a particular strategy under a particular state of nature.

(*v*) ***Opportunity loss table.*** An opportunity loss is the loss incurred because of failure to take the best possible action. Opportunity losses are calculated separately for each state of nature that might occur. Given the occurence of a specific state of nature we can determine the best possible action. For a given state of nature, the opportunity loss of an act is the difference between the payoff of that act, and the payoff of the best act that could have been selected.

In any decision problem, the decision-maker has to choose from the available alternative courses of action the one that yields the best result. If the consequences of each choice are known with certainty, the decision-maker can easily make decisions. But in most of real life problems, the decision-maker has to deal with situations where uncertainty of the outcomes prevails.

3.2.2 Decision-Making Under Certainty

In this case, the decision-maker knows with certainty the consequences of every alternative or decision choice. The decision-maker presumes that only one state of nature is relevant for his purpose. He identifies this state of nature, takes it for granted and presumes complete knowledge as to its occurrence.

3.3 DECISIONS UNDER CONDITIONS OF UNCERTAINTY

When the decision-maker faces multiple states of nature but he has no means to arrive at probability values to the likelihood of occurrence of these states of nature, the problem is a decision problem under uncertainty. Such situations arise when a new product is introduced in the market or a new plant is set up. In business, there are many problems of this 'nature'. Here, the choice of decision largely depends on the personality of the decision-maker.

The following methods are available to the decision-maker in situations of uncertainty.

(*i*) Maximax criterion
(*ii*) Minimax criterion
(*iii*) Maximin criterion
(*iv*) Laplace criterion (Criterion of equally likelihood)
(*v*) Hurwicz alpha criterion (Criterion of realism)

(i) Maximax criterion: The term 'maximax' is an abbreviation of the phrase maximum of the maximums and an adventurous and aggressive decision-maker may choose to take the action that would result in the maximum payoff possible. Suppose for each action there are three possible pay-offs corresponding to three states of nature as given in the following decision matrix:

State of Nature	***Decisions***		
	A_1	A_2	A_3
S_1	220	180	100
S_2	160	190	180
S_3	140	170	200

Maximum under each decision are (220, 190, 200). The maximum of these three maximums is 220. Consequently, according to the maximax criteria the decision to be adopted is A_1.

(ii) Minimax criterion: Minimax is just the opposite of maximax. Application of the minimax criteria requires a table of losses instead of gains. The losses are the costs to be incurred or the damages to be suffered for each of the alternative actions and states of nature. The minimax rule minimizes the maximum possible loss for each course of action. The term 'minimax' is an abbreviation of the phrase minimum of the maximum.Under each of the various actions, there is a maximum loss and the action that is associated with the minimum of the various maximum losses is the action to be taken according to the minimax criterion. Suppose the loss table is:

State of Nature	***Decisions***		
	A_1	A_2	A_3
S_1	0	4	10
S_2	3	0	6
S_3	18	14	0

It shows that the maximum losses incurred by the various decisions are,

A_1	A_2	A_3
18	14	10

Also, the minimum among three maximums is 10 which is under action A_3. Thus according to minimax criterion, the decision-maker should take action A_3.

(iii) Maximin criterion: The maximin criterion of decision-making stands for choice between alternative courses of action assuming pessimistic view of nature. Taking each act in turn, we note the worst possible results in terms of payoff and select the act which maximizes the minimum payoff. Suppose the payoff table is:

State of Nature	*Action*		
	A_1	A_2	A_3
S_1	–80	–60	–20
S_2	–30	–10	–2
S_3	30	15	7
S_4	75	80	25

Minimum under each decision on A_1, A_2 and A_3 are –80 –60 –20 respectively for S_1 state of nature.

The action A_3 is to be taken according to this criterion because it is the maximum among minimums.

(iv) Laplace criterion: As the decision-maker has no information about the probability of occurrence of various events, the decision-maker makes a simple assumption that each probability is equally likely. The expected payoff is worked out on the basis of these probabilities. The act having maximum expected payoff is selected.

Example 3.1: Calculate the maximum expected payoff and the optimal act from the following data:

Events	*Act*		
	A_1	A_2	A_3
E_1	20	12	25
E_2	25	15	30
E_3	30	20	22

Solution: We associate equal probability for each event, say 1/3. Expected pay-offs are:

$$A_1 \rightarrow \left(20\times\frac{1}{3}\right)+\left(25\times\frac{1}{3}\right)+\left(30\times\frac{1}{3}\right)=\frac{75}{3}=25$$

$$A_2 \rightarrow \left(12\times\frac{1}{3}\right)+\left(15\times\frac{1}{3}\right)+\left(20\times\frac{1}{3}\right)=\frac{47}{3}=15.67$$

$$A_3 \rightarrow \left(25\times\frac{1}{3}\right)+\left(30\times\frac{1}{3}\right)+\left(22\times\frac{1}{3}\right)=\frac{77}{3}=25.67$$

Since A_3 has the maximum expected payoff, A_3 is the optimal act.

(v) *Hurwicz alpha criterion*: This method is a combination of maximin criterion and maximax criterion. In this method, the decision-maker's degree of optimism is represented by α, the coefficient of optimism. α varies between 0 and 1. When $\alpha = 0$, there is total pessimism and when $\alpha = 1$, there is total optimism.

Here, the criteria D_1, D_2, D_3, etc., is calculated which is connected with all strategies (state of act) and $D_i = \alpha M_i + (1 - \alpha)\, m_i$ where M_i is the maximum payoff of 'i'th strategy and m_i is the minimum payoff of 'i'th strategy. The strategy with highest value of D_1, D_2, ..., is chosen. The decision-maker will specify the value of α depending upon his level of optimism.

Example 3.2: Calculate the maximum and minimum payoff from the given data to specify the value of α and the act to be selected. Take $\alpha = 0.6$.

NOTES

Events	*Act*		
	A_1	A_2	A_3
E_1	20	12	25
E_2	25	15	30
E_3	30	20	22

Solution: Given is $\alpha = 0.6$

For A_1 maximum payoff = 30

Minimum payoff= 20

$\therefore D_1 = (0.6 \times 30) + (1 - 0.6)20 = 26$

Similarly, $D_2 = (0.6 \times 20) + (1 - 0.6)12 = 16.8$

$D_3 = (0.6 \times 30) + (1 - 0.6)22 = 26.8$

Since D_3 is maximum, select the act A_3.

3.4 DECISION-MAKING UNDER RISK

In this situation, the decision-maker has to face several states of nature. But, he has some knowledge or experience which will enable him to assign probability to the occurence of each state of nature. The objective is to optimize the expected profit, or to minimize the opportuiniy loss.

For decision problems under risk, the most popular methods used are EMV (expected monetary value) criterion, EOL (expected opportunity loss), criterion or EVPI (expected value of perfect information).

(i) Expected monetary value (EMV): When probabilities can be assigned to the various states of nature, it is possible to calculate the statistical expectation of gain for each course of action.

The conditional value of each event in the payoff table is multiplied by its probability and the product is summed up. The resulting number is the EMV for the act. The decision-maker then selects from the available alternative actions, the action that leads to the maximum expected gain (that is the action with highest EMV). Consider the following example. Let the states of nature (Events) be S_1 and S_2 and the alternative strategies (Act) be A_1 and A_2, then the payoff table be as follows:

Events	*Act*	
	A_1	A_2
S_1	30	20
S_3	35	30

Let the probabilities for the states of nature S_1 and S_2 be respectively 0.6 and 0.4.

Then,

EMV for $A_1 = (30 \times 0.6) + (35 \times 0.4) = 18 + 14 = 32$

EMV for $A_2 = (20 \times 0.6) + (30 \times 0.4) = 12 + 12 = 24$

∴ EMV for A_1 is greater.

∴ The decision-maker will choose the strategy A_1.

(ii) Expected opportunity loss (EOL): The difference between the greater payoff and the actual payoff is known as opportunity loss. Under this criterion, the strategy which has minimum expected opportunity loss (EOL) is chosen. The calculation of EOL is similar to that of EMV.

Consider the following example on opportunity loss table. Here A_1 and A_2 are the strategies and S_1 and S_2 are the states of nature.

	A_1	A_2
S_1	0	10
S_2	2	–5

Let the probabilities for two states be 0.6 and 0.4.

EOL for $A_1 = (0 \times 0.6) + (2 \times 0.4) = 0.8$

EOL for $A_2 = (10 \times 0.6) + (-5 \times 0.4) = 6 - 2 = 4$

EOL for A_1 is the least. Therefore, the strategy A_1 may be chosen.

(iii) Expected value of perfect information (EVPI): The expected value of perfect information is the average (expected) return in the long run, if we have perfect information before a decision is to be made.

In order to calculate EVPI, we choose the best alternative with the probability of their state of nature. The expected value of perfect information (EVPI) is the expected outcome with perfect information minus the outcome with maximum EMV.

∴ EVPI = Expected value with perfect information – Maximum EMV

Consider the following example.

Example 3.3: A_1, A_2, A_3 are the acts and S_1, S_2, S_3 are the states of nature. Also known that $P(S_1) = 0.5$, $P(S_2) = 0.4$ and $P(S_3) = 0.1$. Calculate the expected value of perfect information.

Solution: The following is the payoff table:

State of Nature	***Payoff Table***		
	A_1	A_2	A_3
S_1	30	25	22
S_2	20	35	20
S_3	40	30	35

$$\text{EMV for } A_1 = (0.5 \times 30) + (0.4 \times 20) + (0.1 \times 40) = 15 + 8 + 4 = 27$$

$$\text{EMV for } A_2 = (0.5 \times 25) + (0.4 \times 35) + (0.1 \times 30) = 12.5 + 14 + 3 = 29.5$$

$$\text{EMV for } A_3 = (0.5 \times 22) + (0.4 \times 20) + (0.1 \times 35) = 11 + 8 + 3.5 = 22.5$$

The highest EMV is for the strategy A_2 and it is 29.5.

Now to find EVPI, work out the expected value for maximum payoff under all states of nature.

NOTES

	Max. Profit of Each State	Probability	Expected Value (= Prob. × Profit)
S_1	30	0.5	15
S_2	35	0.4	14
S_3	40	0.1	4

∴ Expected payoff with perfect information = 33

∴ Thus, the expected value of perfect information (EVPI) = Expected value with perfect information – Maximum EMV = 33 – 29.5 = 3.5

Example 3.4: You are given the following payoffs of three acts A_1, A_2 and A_3, and the states of nature S_1, S_2 and S_3.

		Act	
State of Nature	A_1	A_2	A_3
S_1	25	–10	–125
S_2	400	440	400
S_3	650	740	750

The probabilities of these states of nature are respectively, 0.1, 0.7 and 0.2. Calculate and tabulate the EMV and conclude which of the acts can be chosen as the best?

Solution:

Act A_1 Prob. × Payoff	Act A_2 Prob. × Payoff	Act A_3 Prob. × Payoff
0.1 × 25 = 2.5	0.1 × –10 = –1	0.1 × –125 = –12.5
0.7 × 400 = 280	0.7 × 440 = 308	0.7 × 400 = 280
0.2 × 650 = 130	0.2 × 740 = 148	0.2 × 750 = 150

∴ EMV for A_1 = 412.5, EMV for A_2 = 455, EMV for A_3 = 417.5

Since EMV is maximum for A_2 choose the Act A_2.

Example 3.5: A management is faced with the problem of choosing one of the products for manufacturing. The probability matrix after market research for the two products was as follows:

		State of Nature	
Act	*Good*	*Fair*	*Poor*
Product A	0.75	0.15	0.10
Product B	0.60	0.30	0.10

The profit that the management can make for different levels of market acceptability of the products are as follows:

State of Nature
Profit (in ₹) for Market

Act	Good	Fair	Poor
Product A	35,000	15,000	5,000
Product B	50,000	20,000	Loss of 3,000

NOTES

Calculate the expected value of the choice of alternatives and advise the management.

Solution: Let us put this information in a payoff matrix with probabilities associated with the states of nature.

State of Nature	*Product A* *Profit × Probability*	*Product B* *Profit × Probability*
Good	35000 × 0.75 = 26250	50000 × 0.60 = 30000
Fair	15000 × 0.15 = 2250	20000 × 0.30 = 6000
Poor	5000 × 0.10 = 500	–3000 × 0.10 = –300
EMV	29000	35700

Since the expected payoff (EMV) of product *B* is greater, product *B* should be preferred by the management.

3.5 MINIMAX REGRET CRITERION

The minimax regret criterion focuses on avoiding regrets that may result from making a non-optimal decision. It is an approach to make decision under uncertainty in which the opportunity cost which is termed as regret, associated with each possible course of action is measured and selects the activity that minimizes the maximum regret or loss is selected by the decision-maker. Regret is the difference between the best and worst possible payoff for each option.

Regret theory says that people foresee regret if they make a wrong choice and take this foresight into consideration when making future decisions. Fear of regret can play a large role in dissuading or motivating someone to do something. For example, in investing, the fear of regret can make investors either risk averse or motivate them to take greater risks. Suppose that an investor buys stock in a growth company based only on a friend's recommendation. After six months, the stock falls to 50 per cent of the purchase price, so the investor sells the stock at a loss. To avoid this regret in the future, the investor will enquire and research any stocks that his friend recommends or he might give it another careless shot depending on his personal nature. On the contrary, suppose the investor did not take the friend's recommendation to buy the stock, but the price increased by 50 per cent instead of decreasing. Thus, to avoid the regret of missing out, the investor will be less risk averse and buy any stocks that his friend recommends in the future or he might not let this incident effect his decision in future. So experience and the personal ability to take risk play a major role in regret theory.

The minimax regret approach is to minimize the worst-case regret. The basic aim is to perform as closely as possible to the optimal course. The minimax criterion applied here is to the regret or difference or ratio of the payoffs rather that to the payoff itself. One benefit of minimax (as opposed to expected regret) is that it is independent of the probabilities of the various outcomes. Thus if regret can be accurately computed, one can reliably use minimax regret. Though, probabilities of outcomes are hard to estimate.

Suppose an investor has to choose between investing in stocks, bonds or the money market and the total return depends on what happens to interest rates. The following table shows some possible returns:

Return	Interest Rates Rise	Static Rates	Interest Rates Fall	Worst Return
Stocks	4	4	12	4
Bonds	2	3	8	2
Money Market	3	2	1	1
Best Return	3	4	12	

The crude minimax choice based on returns would be to invest in the money market, ensuring a return of at least 1. However, if interest rates fell then the regret associated with this choice would be large. This would be –11, which is the difference between the 1 received and the 12 which could have been received if the outturn had been known in advance. A mixed portfolio of about 11.1% in stocks and 88.9% in the money market would have ensured a return of at least 2.22, but if interest rates fell then there would be a regret of about –9.78.

For this example, the following regret table is constructed by subtracting best returns from actual returns:

Regret	Interest Rates Rise	Static Rates	Interest Rates Fall	Worst Regret
Stocks	–7	0	0	–7
Bonds	–5	–1	–4	–5
Money Market	0	–2	–11	–11

Therefore, using a minimax choice based on regret, the best course would be to invest in bonds, ensuring a regret of no worse than – 5. A mixed investment portfolio would do even better, such as 61.1% invested in stocks and 38.9% in the money market would produce regret no worse than about – 4.28.

3.6 PREPARATION OF PAYOFF TABLE

Here, we will study how to prepare payoff table with the help of examples.

Example 3.6: All ink manufacturers produce a certain type of ink at a total average cost of ₹ 3 per bottle and sells at a price of ₹ 5 per bottle. The ink is produced over the weekend and is sold during the following week. According to the past experience, the weekly demand has never been less than 78 or greater than 80 bottles.

You are required to formulate the payoff table.

Solution: The different states of nature are the demand for 78 units, 79 units or 80 units termed as S_1, S_2, S_3.

The alternative courses of action are selling 78 units, 79 units or 80 units termed as A_1, A_2, A_3.

Selling price of ink = ₹ 5 per bottle

Cost price = ₹ 3 per bottle

Calculation of payoffs (Payoff stands for the gain)

Sale quantity × Price – Production Quantity × Cost

$$A_1S_1 = 78\times5 - 78\times3 = 390 - 234 = 156$$
$$A_2S_1 = 78\times5 - 79\times3 = 390 - 237 = 153$$
$$A_3S_1 = 78\times5 - 80\times3 = 390 - 240 = 150$$
$$A_1S_2 = 78\times5 - 78\times3 = 390 - 234 = 156$$
$$A_2S_2 = 79\times5 - 79\times3 = 395 - 237 = 158$$
$$A_3S_2 = 78\times5 - 80\times3 = 395 - 240 = 155$$
$$A_1S_3 = 78\times5 - 78\times3 = 390 - 234 = 156$$
$$A_2S_3 = 79\times5 - 79\times3 = 395 - 237 = 158$$
$$A_3S_3 = 80\times5 - 80\times3 = 400 - 240 = 160$$

(*Explanation:* A_1S_1 means selling quantity is 78 and manufacturing quantity is 78. A_2S_1 means sales 78, production 79, and so on.)

Payoff Table

State of Nature (Events)	*Act (Strategy)*		
	A_1	A_2	A_3
S_1	156	153	150
S_2	156	158	155
S_3	156	158	160

Note: We shall show the state of nature in rows and act in columns.

3.6.1 Preparation of Loss Table

Example 3.7: A small ink manufacturer produces a certain type of ink at a total average cost of ₹ 3 per bottle and sells at a price of ₹ 5 per bottle. The ink is produced over the weekend and is sold during the following week. According to the past experience, the weekly demand has never been less than 78 or greater than 80 bottles in his place.

You are required to formulate the loss table.

Solution: Calculation of regret (opportunity loss)

$A_1S_1 = 0$ (since production and sales are of equal quantities, say 78)

$A_2S_1 = 1\times 3 = 3$ (since one unit of production is in excess whose cost = ₹ 3)

$A_3S_1 = 2\times 3 = 6$ (since 2 units of production are in excess whose unit cost is @ ₹ 3)

$A_1S_2 = 1\times 2 = 2$ (since the demand of one unit is more than produced, the profit for one unit is ₹ 2)

Similarly, $A_2S_2 = 0$ (since Units of production = Units of demand)

$A_3S_2 = 1 \times 3 = 3$ $\quad$ $A_1S_3 = 2\times 2 = 4$

$A_2S_3 = 2\times 1 = 2$ $\quad$ and $\quad$ $A_3S_3 = 0$

NOTES

Opportunity Loss Table

State of Nature (Events)	Action		
	A_1	A_2	A_3
S_1	0	3	6
S_2	2	0	3
S_3	4	2	0

From the payoff table also oportunity loss table can be prepared.

Method: Let every row of the payoff table represent a state of nature and every column represent a course of action. Then, from each row select the highest payoff and subtract all payoffs of that row from it. They are the opportunity losses.

See the following examples.

Example 3.8: The following is a payoff table. From it form a regret (opportunity loss) table.

State of Nature	Pay-off Table		
	A_1	A_2	A_3
E_1	156	153	150
E_2	156	158	155
E_3	156	158	160

Solution:

Opportunity Loss Table

State of Nature	Action		
	A_1	A_2	A_3
E_1	$156 - 156 = 0$	$156 - 153 = 3$	$156 - 150 = 6$
E_2	$158 - 156 = 2$	$158 - 158 = 0$	$158 - 155 = 3$
E_3	$160 - 156 = 4$	$160 - 158 = 2$	$160 - 160 = 0$

Example 3.9: The following is a payoff table.

Action	Events (State of Nature)			
	E_1	E_2	E_3	E_4
A_1	50	300	–150	50
A_2	400	0	100	0
A_3	–50	200	0	100
A_4	0	300	300	0

Suppose that the probabilities of the events in this table are $P(E_1) = 0.15$; $P(E_2) = 0.45$; $P(E_3) = 0.25$; $P(E_4) = 0.15$.

Calculate the expected payoff. Prepare the opportunity loss table (Regret table) and calculate the expected loss of each action.

Solution:

Hint: Rewrite the question with E_1, E_2, E_3, E_4 as rows and A_1, A_2, A_3, A_4 as columns.

Events	*Acts*			
	A_1	A_2	A_3	A_4
E_1	50	400	– 50	0
E_2	300	0	200	300
E_3	–150	100	0	300
E_4	50	0	100	0

Calculation of Expected Payoff (EMV)

A_1 *Pay-off × Prob.*	A_2 *Pay-off × Prob.*	A_3 *Pay-off × Prob.*	A_4 *Pay-off × Prob.*
50 × 0.15 = 7.5	400 × 0.15 = 60	– 50 × 0.15 = –7.5	0 × 0.15 = 0
300 × 0.45 = 135	0 × 0.45 = 0	200 × 0.45 = 90	300 × 0.45 = 135
–150 × 0.25 = –37.5	100 × 0.25 = 25	0 × 0.25 = 0	300 × 0.25 = 75
50 × 0.15 = 7.5	0 × 0.15 = 0	100 × 0.15 = 15	0 × 0.15 = 0
EMV = 112.5	EMV = 85	EMV = 97.5	EMV = 210

Opportunity Loss Table

	A_1	A_2	A_3	A_4
E_1	400 – 50 = 350	400 – 400 = 0	400 + 50 = 450	400 – 0 = 400
E_2	300 – 300 = 0	300 – 0 = 300	300 – 200 = 100	300 – 300 = 0
E_3	300 + 150 = 450	300 – 100 = 200	300 – 0 = 300	300 – 300 = 0
E_4	100 – 50 = 50	100 – 0 = 100	100 – 100 = 0	100 – 0 = 100

Calculation Expected Loss (EOL)

A_1 *Loss × Prob.*	A_2 *Loss × Prob.*	A_3 *Loss × Prob.*	A_4 *Loss × Prob.*
350 × 0.15 = 52.5	0 × 0.15 = 0	450 × 0.15 = 67.5	400 × 0.15 = 60
0 × 0.45 = 0	300 × 0.45 = 135	100 × 0.45 = 45	0 × 0.45 = 0
450 × 0.25 = 112.5	200 × 0.25 = 50	300 × 0.25 = 75	0 × 0.25 = 0
50 × 0.15 = 7.5	100 × 0.15 = 15	0 × 0.15 = 0	100 × 0.15 = 15
EOL = 172.5	EOL = 200	EOL = 187.5	EOL = 75

Example 3.10: A newspaper boy has the following probability of selling a magazine:

No. of Copies Sold	***Probability***
10	0.10
11	0.15
12	0.20
13	0.25
14	0.30

The cost of a copy is 30 paise and the sale price is 50 paise. He cannot return unsold copies. How many copies should he order?

Solution: We can apply either EMV criterion or EOL criterion. Let us apply EMV criterion for which we have to calculate the payoff. Number of copies ordered are the different courses of action. The copies ordered may be 10, 11, 12, 13, 14. Denote them as A_1, A_2, A_3, A_4, A_5.

NOTES

Similarly, the number of copies demanded may be 10, 11, 12, 13 or 14. These demands may be D_1, D_2, D_3, D_4, D_5 and are termed as events. The payoff values are calculated as follows:

Selling price of each item = 50 paise and cost of a copy = 30 paise.

$$A_1D_1 = (10 \times 50) - (10 \times 30) = 200$$
$$A_2D_1 = (10 \times 50) - (11 \times 30) = 170$$
$$A_3D_1 = (10 \times 50) - (12 \times 30) = 140$$
$$A_4D_1 = (10 \times 50) - (13 \times 30) = 110, \text{ etc.}$$

The 25 payoff values can thus be calculated. They are shown as follows:

	A_1	A_2	A_3	A_4	A_5
D_1	200	170	140	110	80
D_2	200	220	190	160	130
D_3	200	220	240	210	180
D_4	200	220	240	260	230
D_5	200	220	240	260	280

The given probabilities are 0.10, 0.15, 0.20, 0.25, 0.30

Calculation of EMV for all the acts.

A_1 *Payoff* × *Prob.*	A_2 *Payoff* × *Prob.*	A_3 *Payoff* × *Prob.*
0.10 × 200 = 20	0.10 × 170 = 17	0.10 × 140 = 14
0.15 × 200 = 30	0.15 × 220 = 33	0.15 × 190 = 28.5
0.20 × 200 = 40	0.20 × 220 = 44	0.20 × 240 = 48
0.25 × 200 = 50	0.25 × 220 = 55	0.25 × 240 = 60
0.30 × 200 = 60	0.30 × 220 = 66	0.30 × 240 = 72
200	215	222.5

A_4 *Payoff* × *Prob.*	A_5 *Payoff* × *Prob.*
0.10 × 110 = 11	0.10 × 80 = 8
0.15 × 160 = 24	0.15 × 130 = 19.5
0.20 × 210 = 42	0.20 × 180 = 36
0.25 × 260 = 65	0.25 × 230 = 57.5
0.30 × 260 = 78	0.30 × 280 = 84
220	205

∴ EMV for the acts A_1, A_2, A_3, A_4 and A_5 are respectively 200, 215, 222.5, 220 and 205.

EMV for A_3 is greater and therefore, A_3 is the optimal act.

∴ No. of copies to be ordered = 12.

Example 3.11: A grocery store with a bakery department is faced with the problems of how many cakes to buy in order to meet a day's demand. The grocer prefers not to sell day-old goods in competition with fresh products. Leftover cakes are, therefore, a complete loss. On the other hand, if a customer desires a cake and all of them have been sold, the disappointed customer will buy elsewhere and the sales will be lost. The grocer has therefore collected information on the past sales or a selected 100-day period as shown in the given table:

Sales Per Day	No. of Days	Probability
25	10	0.10
26	30	0.30
27	50	0.50
28	10	0.10
	100	1.00

Construct the payoff table and the opportunity loss table. What is the optimal number of cakes that should be bought each day? Apply both EMV and EOL criteria. Also, find (and interpret) EVPI (expected value of perfect information). A cake costs ₹ 0.80 and sells for Re 1.

Solution: Let A_1, A_2, A_3 and A_4 be the strategies and S_1, S_2, S_3, S_4 be the different states of nature.

Then A_1, A_2, A_3, A_4 respectively stand for stocking 25, 26, 27, 28 cakes. S_1, S_2, S_3, S_4 respectively stand for demands for 25, 26, 27, 28 cakes.

The conditional payoff values can be obtained as explained in Example 2.3. The payoff values thus obtained are given as follows.

Payoff Table

State of Nature (Demand)	Alternative Strategies			
	$A_1(25)$	$A_2(26)$	$A_3(27)$	$A_4(28)$
$S_1(25)$	5.00	4.20	3.40	2.60
$S_2(26)$	5.00	5.20	4.40	3.60
$S_3(27)$	5.00	5.20	5.40	4.60
$S_4(28)$	5.00	5.20	5.40	5.60

Probabilities are 0.10, 0.30, 0.50, 0.10

EMV for act $A_1 = (5\times0.10)+(5\times0.30)+(5\times0.50)+(5\times0.10)=5.0$

Similarly, EMV values A_2, A_3, A_4 are 5.1, 4.9, 4.2 respectively.

The maximum EMV is found to be strategy A_2. Thus, according to the EMV decision criterion, the store would stock 26 cakes (A_2).

Regret (Opportunity Loss) Table

State of Nature (Demand)	Alternative Strategies			
	$A_1(25)$	$A_2(26)$	$A_3(27)$	$A_4(28)$
$S_1(25)$	0	0.8	1.6	2.4
$S_2(26)$	0.2	0	0.8	1.6
$S_3(27)$	0.4	0.2	0	0.8
$S_4(28)$	0.6	0.4	0.2	0

Probabilities are 0.10, 0.30, 0.50, 0.10

Expected opportunity loss for A_1,

$$= (0\times0.10)+(0.2\times0.30)+(0.4\times0.50)+(0.6\times0.10)=0.32$$

Similarly, for A_2, A_3, A_4 expected opportunity losses are 0.22, 0.42 and 1.12 respectively.

EOL is least for A_2. A_2 is the optimal act.

Hence, 26 is the optimal number to buy.

To find EVPI, select highest payoff in each row and find the expected value. Then we have,

Payoff	*Prob.*	*Payoff × Prob.*
5.00	0.10	0.50
5.20	0.30	1.56
5.40	0.50	2.70
5.60	0.10	0.56
		5.32

∴ Expected value with perfect information = 5.32

Maximum EMV is for Act A_2 which is equal to 5.10.

Expected value of perfect information (EVPI) = Expected value with perfect information – Highest EMV (EMV of A_2) = 5.32 – 5.10 = 0.22.

Example 3.12: A factory produces three varieties of fountain pens. The fixed and variable costs are given as follows:

	Fixed Cost	*Variable Cost*
Type 1	₹2,00,000	₹ 10
Type 2	₹3,20,000	₹ 8
Type 3	₹6,00,000	₹ 6

The likely demands under three situations are given as follows:

Demand	*Units*
Poor	25,000
Moderate	1,00,000
High	1,50,000

If the price of each type is ₹ 20, prepare the payoff table after showing necessary calculations.

Solution: Let T_1, T_2, T_3 stand for *Type 1*, *Type 2*, and *Type 3* and D_1, D_2 and D_3 for poor demand, moderate demand and high demand respectively.

The Payoff (in thousands) = Sales revenue from the estimated demand – Total variable cost – Fixed cost.

$$T_1D_1 = (20\times 25) - (10\times 25) - 200 = 500 - 250 - 200 = +50$$
$$T_2D_1 = (20\times 25) - (8\times 25) - 320 = 500 - 200 - 320 = -20$$
$$T_3D_1 = (20\times 25) - (6\times 25) - 600 = 500 - 150 - 600 = -250$$
$$T_1D_2 = (20\times 100) - (10\times 100) - 200 = 2000 - 1000 - 200 = +800$$
$$T_2D_2 = (20\times 100) - (8\times 100) - 320 = 2000 - 800 - 320 = +880$$
$$T_3D_2 = (20\times 100) - (6\times 100) - 600 = 2000 - 600 - 600 = +800$$
$$T_1D_3 = (20\times 150) - (10\times 150) - 200 = 3000 - 1500 - 200 = +1300$$
$$T_2D_3 = (20\times 150) - (8\times 150) - 320 = 3000 - 1200 - 320 = +1480$$
$$T_3D_3 = (20\times 150) - (6\times 150) - 600 = 3000 - 900 - 600 = +1500$$

Payoff Table (in '000s)

	T_1	T_2	T_3
D_1	50	–20	–250
D_2	800	880	800
D_3	1300	1480	1500

Notes:

1. Expected value of sales = Sum of the products of various values of sales and probabilities.
2. Expected monetary value of an act = Sum of the products of various values of the payoff of the act and probabilities.
3. Expected value of cost = Sum of the products of the various values of the cost and probabilities.
4. Expected value of loss of an act = Sum of the products of the losses of the acts and probabilities.

Example 3.13: A food products company is counterplanning the introduction of a revolutionary new product with new packing to replace the existing product at much higher price (S_1) or a moderate change in the composition of the existing product with a new packaging at a small increase in price (S_2) or a small change in the composition of the existing product except the word, 'New with a negligible increase in the price (S_3). The three possible states of nature of events are (*i*) High increase in sales (N_1), (*ii*) No change in sales (N_2) and (*iii*) Decrease in sales (N_3). The marketing department of the company worked out the pay-offs in terms of yearly new profits for each of the strategies or these events (expected sales). This is represented in the following table.

Strategies	***Payoffs State of Nature***		
	N_1	N_2	N_3
S_1	700	300	150
S_2	500	450	0
S_3	300	300	300

Which strategy should the executive concerned choose on the basis of: (*i*) Maximin criterion, (*ii*) Maximax criterion, (*iii*) Minimax regret criterion and (*iv*) Laplace criterion?

Solution: Writing the payoff table properly,

State of Nature	***Strategies***		
	S_1	S_2	S_3
N_1	700	500	300
N_2	300	450	300
N_3	150	0	300

(*i*) Maximin criterion

	Minimum Payoffs
S_1	150
S_2	0
S_3	300

Maximum of these minima = 300

NOTES

The executive should choose strategy S_3.

(*ii*) Maximax criterion

	Maximum Payoffs
S_1	700
S_2	500
S_3	300

Maximum of maxima = 700

∴ The executive can choose strategy S_1.

(*iii*) Minimax regret criterion

Opportunity Loss Table

	S_1	S_2	S_3
N_1	700 – 700 = 0	700 – 500 = 200	700 – 300 = 400
N_2	450 – 300 = 150	450 – 450 = 0	450 – 300 = 150
N_3	300 – 150 = 150	300 – 0 = 300	300 – 300 = 0

Maximum	*Opportunity Loss*
For S_1	– 150
For S_2	– 300
For S_3	– 400

The executive should choose strategy S_1 since it minimizes the maximum of the losses.

(*iv*) Laplace criterion

Assigning equal probability (say 1/3) to each state of nature, calculate the expected monetary value for each strategy.

Prob. × Payoff:

$$\left(\frac{1}{3}\times 700\right)+\left(\frac{1}{3}\times 300\right)+\left(\frac{1}{3}\times 150\right)=\frac{1150}{3}=383.33$$

$$\left(\frac{1}{3}\times 500\right)+\left(\frac{1}{3}\times 450\right)+\left(\frac{1}{3}\times 0\right)=\frac{950}{3}=316.666$$

$$\left(\frac{1}{3}\times 300\right)+\left(\frac{1}{3}\times 300\right)+\left(\frac{1}{3}\times 300\right)=\frac{900}{3}=300$$

Since the EMV is highest for strategy 1, the executive should select strategy S_1.

Check Your Progress

1. Classify decisions.
2. What are the types of decision-making situations?
3. What is Expected Monetary Value (EMV)?
4. What is Expected Opportunity Loss (EOL)?
5. What is Expected Value of Perfect Information (EVPI)?

3.7 TYPES OF DECISION MODELS

NOTES

Once the objective, alternative strategies and the decision-making environments are known, the next step which a decision-maker faces is to select the decision model which can fit into his problem. There are various models used in decision-making. Some of the models are as follows.

3.7.1 Deterministic Decision Model

Deterministic model is related to deterministic situation. Deterministic decision payoffs are the simplest possible payoffs. The objectives and strategies in this model have to be listed and then the payoff for each strategy towards each objective is determined. For example, if there are two objectives O_1 and O_2, the strategies to be selected are S_1 and S_2 and then the related payoffs can be shown in matrix form as under:

Objectives/Strategies	O_1	O_2	*Total Payoff* Σa_{ij}
S_1	a_{11}	a_{12}	Σa_{1j}
S_2	a_{21}	a_{22}	Σa_{2j}

Here a_{ij} ($i = 1, 2; y = 1, 2$) refers to pay-offs of *i*th strategy towards *j*th objective. Total payoff for strategy 1 is Σa_{1j} (i.e., a_{11} payoff towards objective 1 and a_{12} payoff towards objective 2) and for strategy 2 is Σa_{2j}. The optimum strategy would be the one having the largest total payoff (i.e., maximum of Σa_{1j} and Σa_{2j}).

In general, with *m* objectives and *n* strategies the decision payoff is as follows:

Objectives/Strategies	O_1	$O_2 ... O_m$	*Total Payoff*
S_1	a_{11}	$a_{12} ... a_{1m}$	Σa_{1j}
S_2	a_{21}	$a_{22} ... a_{2m}$	Σa_{2j}
.	.	.... .	.
.	.	.... .	.
.	.	.... .	.
S_n	a_{n1}	$a_{n2} ... a_{nm}$	Σa_{nj}

Here Σa_{ij} again refers to payoff of *i*th strategy towards *j*th objective. The optimum strategy in this case would be the one having the largest payoff (i.e., maximum of $\Sigma a_{1j}, \Sigma a_{2j}, \ldots \Sigma a_{nj}$).

The decision-making under certainty situation involves the following steps:

(i) Determine the alternative courses of action.

(ii) Calculate the payoffs, one for each course of action.

(iii) Select the alternative with largest profit or smallest cost either by the method of complete enumeration (if the number of alternatives is small) or with the aid of appropriate mathematical models.

3.7.2 Probabilistic or Stochastic Decision Model

Probabilistic model or what is known as the stochastic decision model is related to risk situation. Risk situation, as has already been explained, is one where there are many states of nature and the decision-maker knows the probability of occurrence of each such state. Decision pay-offs are not fixed but generally happens to be a random variable. Pay-offs are determined partly by chance and partly by the strategies adopted. Hence, in a probabilistic decision model, a decision is made in favour of that strategy which has the maximum expected payoff.

Let us consider a simple example in which we have three objectives with three strategies. Objectives are denoted by O_1, O_2 and O_3 and strategies by S_1, S_2 and S_3. The payoff matrix can be stated as under:

Objectives/Strategies	O_1	O_2	O_3
S_1	a_{11}	a_{12}	a_{13}
S_2	a_{21}	a_{22}	a_{23}
S_3	a_{31}	a_{32}	a_{33}

The matrix of risk function (or probability) can similarly be denoted as under:

Objectives/Strategies	O_1	O_2	O_3
S_1	p_{11}	p_{12}	p_{13}
S_2	p_{21}	p_{22}	p_{23}
S_3	p_{31}	p_{32}	p_{33}

Where *ij* refers to the probability of selecting *i*th strategy towards the achievement of *j*th objective. Also p_{ij} for all *i*, *j* or $\Sigma p_y = 1$.

After knowing the above stated two matrices, the next step is to calculate the expected pay-offs *ij* which can also be termed as Expected Monetary Value (or EMV). ΣE_{ij} is equal to the multiplication of decision payoff elements to the corresponding probabilities. The expected payoff matrix would be as follows:

Objectives/Strategies	O_1	O_2	O_3	Total Expected Payoff (or EMV)
S_1	E_{11}	E_{12}	E_{13}	ΣE_{1j}
S_2	E_{21}	E_{22}	E_{23}	ΣE_{2j}
S_3	E_{31}	E_{32}	E_{33}	ΣE_{3j}

The best strategy in this case would be the one having the largest total expected payoff or the EMV (i.e., maximum of ΣE_{1j}, ΣE_{2j} and ΣE_{3j}). The similar treatment can be extended for *n* strategies and with *m* objectives.

3.7.3 Rules/Techniques for Decision-Making Under Risk Situation

There are several rules and techniques for decision-making under risk situation. Important ones are:

1. Maximum likelihood rule.
2. Expected payoff criterion:
   (*i*) EMV criterion
   (*ii*) EOL criterion
3. Decision trees.

4. Utility functions or the utility curves.
5. Bayesian decision rule (Posterior analysis)

We shall now explain all these decision rules and techniques one by one.

NOTES

1. Maximum Likelihood Rule

Under this rule the decision-maker selects the most likely alternative. For example, if the probability distribution of demand is as follows:

Demand (Units)	0	1	2	3	4	5	6
Probability	0.1	0.1	0.4	0.05	0.05	0.2	0.1

As per the table the most likely demand is a demand of 2 units and if one has to place the order, then he should place for 2 units as per the most likelihood rule. The disadvantage of this rule is that no consideration is given to less likely but more consequential results.

2. Expected Payoff Criterion (EMV and EOL)

Expected monetary value (or EMV) criterion: For a probabilistic decision model the usual criterion is that of Expected Monetary Value. The decision-maker will have to adopt the following steps to work out the EMV:

(*i*) The decision-maker should clearly state all possible actions that he thinks reasonable for the period (or periods) in question and also the possible outcomes of the actions.

(*ii*) The decision-maker must then state the probability distribution concerning each possible action for which purpose he may use either a-priori or empirical methods of calculating probabilities. In simple words, the decision-maker should assign a probability weight to each of the possible actions or the states of nature.

(*iii*) The decision-maker must finally use some yardstick (usually rupees) that measures the value of each outcome. Thus in other words, it means that the decision-maker should compute for each state of nature the consequences of the given act.

(*iv*) He can then calculate the total expected payoff (or expected monetary value) concerning each action and its outcome. This is done by summing up the product of the probability of each state of nature and the consequences of that act.

(*v*) The action and outcome with the highest expected value should be finally selected.

In context of EMV, it should be kept in view that EMV technique is adequate only in those cases where the potential losses are not too great and the perspective profit range is narrow. But in problems which involve large potential losses some other techniques of decision-making are generally adopted.

This is explained with the help of following examples.

Example 3.14: Suppose a businessman wants to decide whether to stock commodity *X* or commodity *Y*. He can stock either but not both. If he stocks *X* and if it is a success, he feels that he can make ₹ 200 but if it is a failure he will lose ₹ 500. If he stocks *Y* and if it is a success he feels that he can make ₹ 400, but if it is a failure he would lose ₹ 300. The question is, Which commodity *X* or *Y* should he stock? He has the following probability distribution in view:

Probability	*With Stock of Commodity X*	*With Stock of Commodity Y*
Success	0.80	0.60
Failure	0.20	0.40

Solution: We can write the payoff matrix (in rupee terms) as follows for the given information:

Objective/Strategy	*Success*	*Failure*
Commodity *X*	+200	–500
Commodity *Y*	+400	–300

The probability matrix being already given in the question, we can write the expected payoff (or EMV) matrix as under:

Objective/Strategy	*Success*	*FailureTotal*	*Expected Payoff (or EMV)*
Commodity *X*	(0.8) (200)	(0.2) (–500)	160 – 100 = +60
Commodity *Y*	(0.6) (400)	(0.4) (–300)	240 – 120 = +120

From the above matrix it is clear that EMV of stocking commodity *X* is (+) 60 rupees and that of stocking commodity *Y is* (+) 120 rupees. Clearly, the businessman should choose to stock commodity *Y*.

Interpretation of EMV

EMV should be correctly interpreted and understood by a decision-maker. In the given problem the expected monetary value is ₹ 120 which does not mean an assured profit of ₹ 120. It is just the expected value of making a profit. It simply means that if the businessman made this decision of stocking commodity *Y* several times, he would on the average make a profit of ₹ 120. But if he stocks commodity *Y*, say just once, he may even lose ₹ 300. The correct conclusion is that the chances of a greater profit are there with the stocking of commodity *Y*.

Example 3.15: Suppose a grocer is faced with a problem of how many cases of milk to stock to meet tomorrow's demand. All the cases of milk left at the end of the day are worthless. Each case of milk is sold for ₹ 8 and is purchased for ₹ 5. Hence, each case sold brings a profit of ₹ 3 but if it is not sold at the end of the day, then it must be discarded resulting in a loss of ₹ 5. The historical record of the number of cases of milk demanded is as follows:

No. of Cases of Milk Demanded	*No. of Times Demanded*	*Probability of Each Event*
0—12	0	0.00
13	5	0.05
14	10	0.10
15	20	0.20
16	30	0.30
17	25	0.25
18	10	0.10
Over 18	0	0.00
Total	100	1.00

NOTES

What should be the optimal decision of the grocer concerning the number of cases of milk to stock? Assuming that the grocer has a perfect knowledge, evaluate what would be his expected profits?

Solution: The situation facing the grocer can be structured in the form of a matrix which shows conditional values (or conditional profits) as follows:

Matrix of Conditional Values (or Profits)

Event: Demand or State of Nature	*Possible Action Concerning Stock Policy*					
	Stock 13	*Stock* 14	*Stock* 15	*Stock* 16	*Stock* 17	*Stock* 18
13	39	34	29	24	19	14
14	39	42	37	32	27	22
15	39	42	45	40	35	30
16	39	42	45	48	43	38
17	39	42	45	48	51	46
18	39	42	45	48	51	54

Demand upto 12 cases and also demand over 18 cases have not been considered in the above matrix for their probabilities are zero on the basis of the past information.

It may be pointed out here that conditional value (or profit) means the actual profit which would result following a given action, conditional upon a given event occurring. Thus, in the given case, for example, if 16 cases of milk are ordered and 14 are sold, then the conditional profit would be ₹ 32 to be worked out as under:

Profit per case of milk sold = ₹ 3

Profit of 14 cases sold = ₹ 42

2 cases remain unsold and are worthless resulting in a loss of ₹ 5 per case.

Total loss on 2 cases = ₹ 10

∴ Conditional profit = (₹ 42 – ₹ 10) = ₹ 32

Conditional profit for all other possible actions can be worked out in a similar manner and a matrix of conditional profits (values) can be prepared.

The best action to be taken (or the optimal decision) is found by calculating the EMV for each stock action and then choosing the highest EMV. This can be done as shown here:

NOTES

Event: Demand	*Probability of Each Event*	*Conditional Profit of Possible Action Stock 13*	*Expected Value of Possible Action Stock 13*
13	0.05	39	1.95
14	0.10	39	3.90
15	0.20	39	7.80
16	0.30	39	11.70
17	0.25	39	9.75
18	0.10	39	3.90

Expected monetary value = 39.00 (of stocking 13 cases of milk)

Similarly, the EMV for all other possible actions of stocking milk cases can be worked out which would be as follows:

Possible Action	*EMV*
Stock 14	0.05(34) + 0.10(42) + 0.20(42) + 0.30(42) + 0.25(42) + 0.10(42) = 41.6
Stock 15	0.05(29) + 0.10(37) + 0.20(45) + 0.30(45) + 0.25(45) + 0.10(45) = 43.4
Stock 16	0.05(24) + 0.10(32) + 0.20(40) + 0.30(48) + 0.25(48) + 0.10(48) = 43.6
Stock 17	0.05(19) + 0.10(27) + 0.20(35) + 0.30(43) + 0.25(51) + 0.10(51) = 41.4
Stock 18	0.05(14) + 0.10(22) + 0.20(30) + 0.30(38) + 0.25(46) + 0.10(54) = 37.2

The highest EMV is 43.6 rupees corresponding to action of stocking 16 cases of milk. Thus, the optimal course of action under the given condition of risk is to stock 16 cases of milk.

3.7.4 Expected Profits with Perfect Knowledge (or Information) and the Expected Value of Perfect Information

Perfect knowledge means that the decision-maker knows demand when he orders the goods. In such a situation, the following conditional values (which can be read from the above stated matrix of conditional values) would be relevant:

Event: Demand	*Conditional Value*	
13	With possible action: Stock 13	₹39
14	With possible action: Stock 14	₹42
15	With possible action: Stock 15	₹45
16	With possible action: Stock 16	₹48
17	With possible action: Stock 17	₹51
18	With possible action: Stock 18	₹54

Expected profits with perfect knowledge now can be worked out as under:

Event: Demand	*Probability Perfect Knowledge*	*Conditional Value with Value*	*Expected*
13	0.05	39	1.95
14	0.10	42	4.20
15	0.20	45	9.00
16	0.30	48	14.40
17	0.25	51	12.75
18	0.10	54	5.40

∴ Expected profits with perfect knowledge = ₹ 47.70 (or EMV with condition of certainty)

∴ EMV with condition of certainty is ₹ 47.70

∴ EMV with condition of risk is ₹ 43.60. Hence, the expected value of perfect information or EVPI = (47.70 – 43.60) = ₹ 4.10 being the value of transforming risk into certainty.

Expected Opportunity Loss (EOL) Criterion: The grocer can also choose the best act in the given problem by minimizing expected opportunity loss. For this purpose a matrix showing conditional opportunity losses can be prepared as follows:

It may be pointed out here that conditional opportunity loss means the relative loss (i.e., the profit not earned) following a given action and conditional upon a given event occurring. Thus, in the given case, for example, if 16 cases of milk are ordered and 14 are sold, the conditional loss would be ₹ 10 (i.e., loss of ₹ 5 per case of unsold milk). If 13 cases are ordered and the demand happens to be of 18 cases, then the grocer would not be able to make a profit of ₹ 15 which he could have made had he ordered for 18 cases of milk. Thus, ₹ 15 is the conditional or opportunity loss in this concerning event for the grocer losses the opportunity to sell 5 additional cases of milk. Conditional opportunity losses for all other possible actions can be worked out in a similar manner and exhibited in the matrix of conditional opportunity losses as given below.

Table 3.1 *Matrix of Conditional Opportunity Losses*

Event: Demand or State of Nature	*Possible Actions Concerning Stock Policy*					
	Stock 13	***Stock 14***	***Stock 15***	***Stock 16***	***Stock 17***	***Stock 18***
13	0	5	10	15	20	25
14	3	0	5	10	15	20
15	6	3	0	5	10	15
16	9	6	3	0	5	10
17	12	9	6	3	0	5
18	15	12	9	6	3	0

The best action to be taken (or the optimal decision) is found by calculating the EOL for each stock action and then choosing the smallest EOL. This can be done as follows:

Event: Demand	*Probability of Each Event*	*Conditional Opportunity Loss of Possible Action*	*Expected Value*
13	0.05	0	0.00
14	0.10	3	0.30
15	0.20	6	1.20
16	0.30	9	2.70
17	0.25	12	3.00
18	0.10	15	1.50

Expected opportunity loss = 8.70 rupees of stocking 13 cases of milk.

Similarly, the EOL for all other possible actions of stocking milk cases can be worked out which would be as given here:

Possible Action	*EOL (Rupees)*
Stock 14	0.05(5) + 0.10(0) + 0.20(3) + 0.30(6) + 0.25(9) + 0.10(12) = 6.1
Stock 15	0.05(10) + 0.10(5) + 0.20(0) + 0.30(3) + 0.25(6) + 0.10(9) = 4.3
Stock 16	0.05(15) + 0.10(10) + 0.20(5) + 0.30(0) + 0.25(3) + 0.10(6) = 4.1
Stock 17	0.05(20) + 0.10(15) + 0.20(10) + 0.30(5) + 0.25(0) + 0.10(3) = 6.3
Stock 18	0.05(25) + 0.10(20) + 0.20(15) + 0.30(10) + 0.25(5)+ 0.10(0) = 10.5

The smallest EOL is ₹ 4.1 corresponding to action of stocking 16 cases of milk which is the optimum action under given condition of risk. This answer is exactly the same as we had worked it out with EMV.

3.7.5 The Effect of Salvage Value

In the above illustration it was assumed that the unsold cases of milk left at the end of the day were completely worthless but in real life such unsold quantity, may have some value known as the *'salvage value'*. The effect of such salvage value is that it reduces the loss from overstocking. Suppose the salvage value of a case of milk remaining unsold is ₹ 2, then the overstock loss would be (₹ 5 – ₹ 2) ₹ 3 per case of milk. The best action would be worked out as per the procedure outlined above keeping in view the over stock loss after making adjustment concerning the salvage value.

3.7.6 Use of Marginal Analysis

In many problems the procedure to find out the best action either through EMV or through EOL would be a tedious one because of the number of computations required. Marginal analysis provides the alternative to this but only in cases where the gains increase (or losses decrease) linearly. In our example, each additional case of milk sold brings a gain of ₹ 3 and each case of milk not sold causes a loss of ₹ 5. Thus, the given question satisfies the linearity assumption.

Marginal analysis starts considering that an additional unit bought will either be sold or it will not be sold. If p represent the probability of selling one additional unit, then $(1 - p)$ must be the probability of not selling the additional unit, for the sum of the probabilities of these two events must be one. If the additional unit purchased is sold, we shall realize an increase in our conditional profit and such an increase is known as the marginal profit (*MP*). But if the additional unit purchased is not sold, then it will reduce our conditional profit and the amount of reduction is known as the marginal loss (*ML*).

Thus the additional units should be stocked (purchased) as long as the expected marginal profit from stocking each of them is greater than the expected marginal loss from stocking each. The size of each order should be increased upto the point where *MP* and *ML* are equal. Thus, we should stock upto the point where expected marginal profit [$p(MP)$] is greater than or equal to the expected marginal loss [$(1-p)(ML)$]. In symbolic form we can state as follows:

Stock upto the point where,

$$p(MP) \geq (1-p)(ML)$$

or,
$$p(MP) \geq ML - pML$$

or,
$$p(MP) + p(ML) \geq ML$$

or,
$$p \geq \frac{ML}{ML + MP}$$

Example 3.16: Solve the problem given in example 3.5 by applying the marginal analysis.

Solution: An additional case of milk stocked (purchased) and sold increases the conditional profit by ₹ 3. Hence marginal profit or *MP* = ₹ 3.

An additional case of milk stocked but not sold causes reduction in the conditional profit by ₹ 5. Hence marginal loss or *ML* = ₹ 5.

Hence, stock upto the point where,

$$p \geq \frac{ML}{ML + MP}$$

or,
$$p \geq \frac{5}{5+3}$$

or,
$$p \geq 0.625$$

The given probability distribution concerning demand is as follows:

Event: Demand	*Probability*
13	0.05
14	0.10
15	0.20
16	0.30
17	0.25
18	0.10

The above table shows that the probability of selling 18 cases of milk is 0.10. The probability of selling 17 cases of milk or more is,

$$(0.25 + 0.10) = 0.35$$

The cumulative probability of selling 16, 15, 14 and 13 cases of milk can be worked out in the same manner and can be put as under:

Demand (Sales) of	*Cumulative Probability Cases of Milk*
13 or more	1.00
14 or more	0.95
15 or more	0.85
16 or more	0.65
17 or more	0.35
18 or more	0.10

Since the cumulative probability of selling 14 or more is 0.95 we should stock 14 cases of milk (because p is greater than 0.625). On similar reasoning we should clearly stock 15 cases of milk. We can as well stock 16 cases of milk (because p is greater than 0.625). But after this we find p only 0.35 for 17 or more and 0.10 for 18 or more which values are less than 0.625. Hence, the optimal action is to stock 16 cases of milk. This result is the same as we had worked out earlier through EMV and EOL techniques.

3.7.7 Competitive Decision Model

Competitive decision model is related to the situation of uncertainty. As pointed out earlier, in a situation of uncertainty, the probabilities of occurrence of the different events (or the states of nature) are not known. Under the situation of uncertainty the decision-maker though recognizes different potential states of nature but cannot confidently estimate the probabilites of their occurrence. This in other words means that the decision-maker has to act with imperfect information in such a situation. Consequently, there is no single best criterion for selecting a strategy to deal with such a situation but there are different criteria available for selecting a strategy. The following criteria deserve mention in this context:

(*i*) *Maximin (or minimax) decision rule.* This rule is also known as the criterion of pessimism. Under this rule the decision-maker is completely pessimistic. He assumes that the situation will always be disadvantageous. On this presumption a new column labeled 'worst' is developed in the given matrix and from this column, the best entry is selected. Thus, under maximin rule the decision-maker selects that strategy which gives largest of the minimum pay-offs, i.e., the maximum of the minimum gains or in case of a cost (or loss) matrix, minimum of the maximum cost or losses. One drawback of this criterion is that the decision is based on only a small portion of the available information which is entered in the column labeled as 'worst' and the rest of that data is ignored.

(*ii*) *Maximax (or minimin) decision rule.* This is also known as criterion of optimism. Under this rule the decision-maker is quite optimistic. He assumes that the situation will always be to his advantage. He, therefore, selects the strategy which yields him the best possible payoff or the best of bests. Under this rule the decision-maker searches for the best possible payoff for each alternative. These are placed in a new column to the right of the given matrix and the alternative with the best payoff in this newly added column is then selected. If it is a profit matrix, the decision-maker selects a strategy which yields him the highest of the maximum pay-offs, i.e., maximum of the maximums. In case of cost matrix, he selects a strategy which yields him the lowest of the minimum pay-offs, i.e., minimum of the minimums.

NOTES

(iii) *Savage decision rule.* This rule is also known as regret decision rule. It is based on general insurance against risk. It ensures against the maximum possible risk. Under it, one adopts the strategy which causes minimum of the maximum possible regrets and the given payoff matrix is converted into a regret matrix. This is done by subtracting each entry in the payoff matrix from the largest entry in its column. The largest entry in a column will have zero regret. Thus, in each cell we enter the difference between what the decision-maker would have done if he had known which outcome would occur, and the choice is represented by the cell. Once the regret matrix is formed, the minimax criterion can be applied to it to select the best course of action.

(iv) *Hurwicz decision rule.* Hurwicz has developed a decision rule based on the maximin and maximax rules with an index of optimism α and an index of pessimism $(1 - \alpha)$. The value of α always lies between zero and one. The decision-maker should assign a value to α somewhere between 0 and 1. The value of α nearer to one means the decision-maker is optimistic, and nearer to zero reflects a pessimistic decision-maker whereas $\alpha = 1/2$ reflects a neutralist. Largest and smallest values, say V, u respectively, be taken for each and every strategy and then the weighted value be determined as under:

$$\text{Weighted value} = \alpha\ V + (1 - \alpha).u$$

Then the strategy having the highest weighted value as per the above formula given by Hurwicz is selected. The major difficulty in applying this rule is the measurement of the value of α. Although more information than in minimax is being used while applying Hurwicz rule, yet only the two extreme pay-offs (viz., V and μ) are considered and the remaining information is ignored. Hurwicz rule is sometimes also referred as Hurwicz- α index.

(v) *Laplace decision rule.* This rule is based on the assumption (in case the probabilities are not known) that the probabilities of different states of nature for a given strategy are all equal (i.e., all states of nature are equally likely to occur). Considering these equal probabilities the expected pay-offs will be calculated and then the strategy with the highest expected payoff is selected. Laplace rule is also known as the criterion of insufficient reason. The major weakness of this criterion is that there is absolutely no reason to assume the probabilities of different states of nature are all equal.

From the above description it is clear that there is no single best rule for decision-making under the situation of uncertainty. There are several models (decision rules) for the purpose which usually point to different selections of alternatives. The choice for the selection of a model should be left to the decision-maker who should ultimately decide as per his own skill and experience considering the environment, firm's policy and other relevant factors. Thus decision-making under uncertainty is relatively more difficult

NOTES

task than it is under the situation of certainty or under risk situation. Normally decision-making under uncertainty should be avoided but if it becomes necessary then enough information to the extent possible should first be acquired.

3.7.8 Limitations and Advantages of Decision Models

Decision-making is a basic function of manager. Though it is a valuable guide to the managers but it has the following limitations and advantages:

The advantages of decision models are as follows:

1. Information technology is developing rapidly and provides policies to decision-makers. It also provides large amounts of information that require processing and analysis.
2. Decision support systems/models (DSS) aim to provide tools that not only help such analyses, but enable the decision-maker to experiment and simulate the effects of different policies and selection strategies. The specific context of this model is administrative decision-making using large educational databases.
3. A decision support system (DSS) developed to enable systematic exploration of the educational database, allows users to identify variables of interest and to actively change the attribute values of these attributes thus revealing the consequences of these decisions on policies and selection processes.
4. Further refinements allow users to introduce conditional rules which take into account other variables required, for example, local circumstances. An active Document Manager, used in conjunction with DSS, allows decision-makers to manipulate the structure of the document in which the decision-making interactions have been placed.

The limitations of decision models are as follows:

1. It can be extremely time-consuming.
2. It has the potential for 'hung' decisions if someone is unwilling to go along with a decision.
3. It helps for decisions already made to be revisited and slow down a planning progress.
4. It can allow for multiple decisions to be on the table at the same time.
5. It can possibly create confusion.

It may be perceived as unfair.

Check Your Progress

6. When is a problem decision problem under certainty?
7. Name the methods which are useful for decision-making under uncertainty.

3.8 DECISION TREE ANALYSIS

NOTES

A decision tree is a graphical representation of various decision alternatives, states of nature, probabilities attached to the states of nature and the conditional benefits as well as losses. In constructing a tree diagram, following two types of nodes are used:

(*i*) Decision node represented by a square

(*ii*) State of nature node (chance node) represented by a circle.

Alternative courses of action (strategies) start from the decision node as main branches. At the end of each main branch, there is a state of nature node from which emerge chance events in the form of sub-branches.

The corresponding payoffs and the associated probabilities with alternative courses as well as the chance events are shown on these branches. At the end of the chance branches, the expected values of the outcome are shown.

The general approach used in decision tree analysis is to work backwards through the tree from right to left, computing the expected value of each chance node. Then select the particular branch having a decision node, which leads to the chance node with the maximum expected value.

Steps in Decision Tree Analysis

1. Identify the decision points and alternatives at each point systematically.
2. At each decision point, determine the probability and the associated payoff with each course of action.
3. Compute the expected payoff (EMV) for each course of action. Start from the extreme right and move towards the left.
4. Choose the course of action that yields the best payoff for each of the decisions. Proceed backward to the next stage of decision points.
5. Repeat the above steps till the first decision point is reached.
6. Identify the best course of action to be adopted from the beginning to the end, under various possible outcomes as a whole.

Example 3.17: A company is going to develop a new product in the market. Three alternative decisions are available for the management.

A_1: Advertising on television, where advertising cost is ₹ 3,000 per day.

A_2: Appointing salesmen for marketing. The cost is ₹ 1,200 per day.

A_3: Conducting an exhibition, where the cost is ₹ 900 per day.

The unit price is fixed at ₹ 25, product and the cost of units associated with the respective decision alternatives are 9, 5 and 11. The expected demand for the product is as follows.

Demand	200	300	400	500
Probability	0.3	0.2	0.4	0.1

NOTES

The company has to decide upon the best alternative among the three decisions.

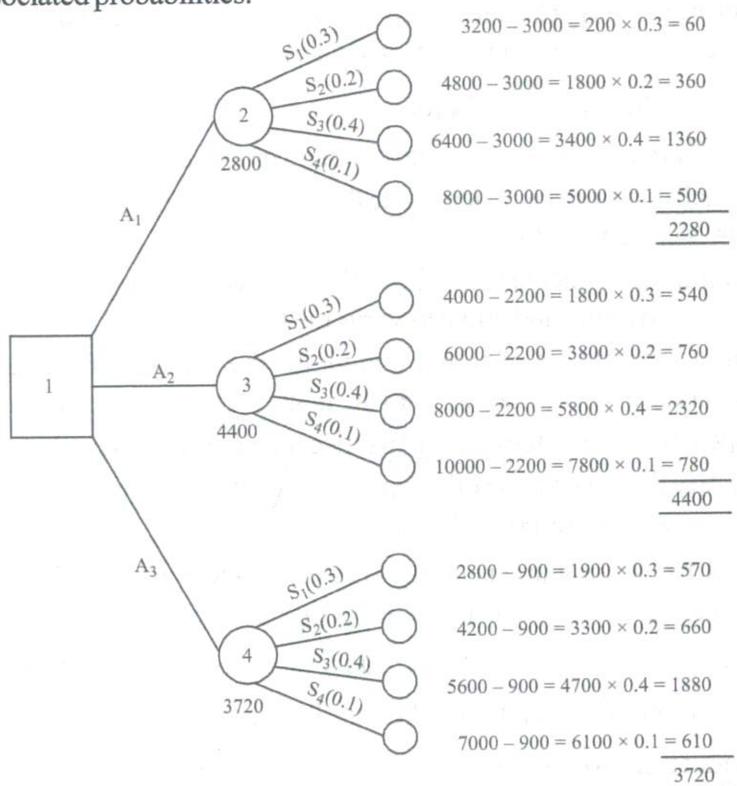
Solution:

1. Formulate the payoff table for the three alternative decisions and the state of nature by using,

$$\text{Profit} = (\text{Fixed price} - \text{Cost of units}) \times \text{Demand}$$

Alternatives	*State of nature (Demand)*			
	200	***300***	***400***	***500***
A_1 (9)	$(25-9) \times 200 = 3200$	$(25-9) \times 300 = 4800$	$(25-9) \times 400 = 6400$	$(25-9) \times 500 = 8000$
A_2 (5)	$(25-5) \times 200 = 4000$	$(25-5) \times 300 = 6000$	$(25-5) \times 400 = 8000$	$(25-5) \times 500 = 10000$
A_3 (11)	$(25-11) \times 200 = 2800$	$(25-11) \times 300 = 4200$	$(25-11) \times 400 = 5600$	$(25-11) \times 500 = 7000$

2. Construct a decision tree by using the above payoff values with their associated probabilities.



Conclusion: Since the maximum payoff for alternative 2 = ₹ 4,400, it is the best alternative and should be selected.

Example 3.18: A manager has two independent investments *A* and *B* available to him, but he lacks the capital to undertake both of them simultaneously. He can choose to take *A* first and then stop, or if *A* is successful only then take *B* or vice versa. The probability of success for *A* is 0.7, while for *B* it is 0.4. Both investments require an initial capital outlay of ₹ 2,000 and both return nothing if the venture is unsuccessful. Successful completion of *A* will return ₹ 3,000 and successful completion of *B* will return ₹ 5,000. Draw the decision tree and determine the best strategy.

Solution:

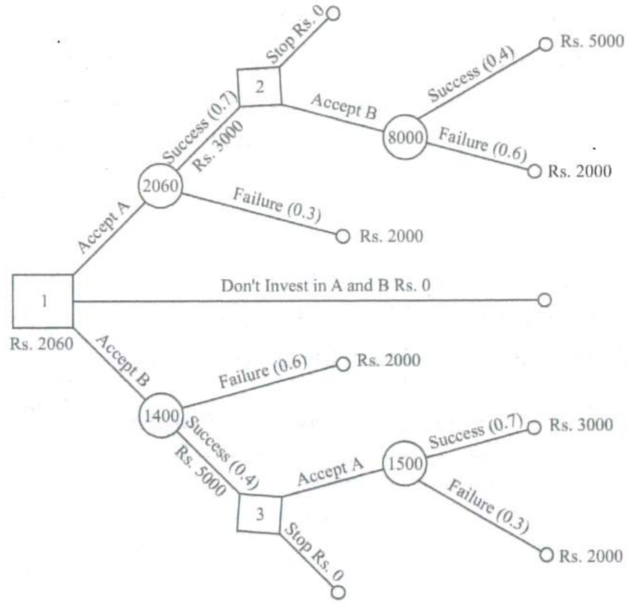
1. Compute the various action-event combinations and tabulate the resulting payoffs of the problem. The net EMV corresponding to various event decision points are indicated in the decision tree.

Decision alternative	*Decision*	*Event*	*Probability*	*Conditional value*	*Expected value*
3	Accept *A*	Success	0.7	3000	2100
		Failure	0.3	(–) 2000	(–) 600
				Net EMV	1500
	Stop	0	0	0	0
2	Accept *B*	Success	0.4	5000	2000
		Failure	0.6	(–) 2000	1200
				Net EMV	(–) 800
	Stop	0	0	0	0
1	Accept *A*	Success	0.7	3800	2660
		Failure	0.3	(–) 2000	(–) 600
					2060
	Accept *B*	Success	0.4	6500	2600
		Failure	0.6	(–) 2000	(–) 1200
				Net EMV	1400
	Stop	0	0	0	0

Conclusion: Since the EMV is maximum in decision 1, the optimal decision is to accept investment *A*, and if it is successful, then accept *B*.

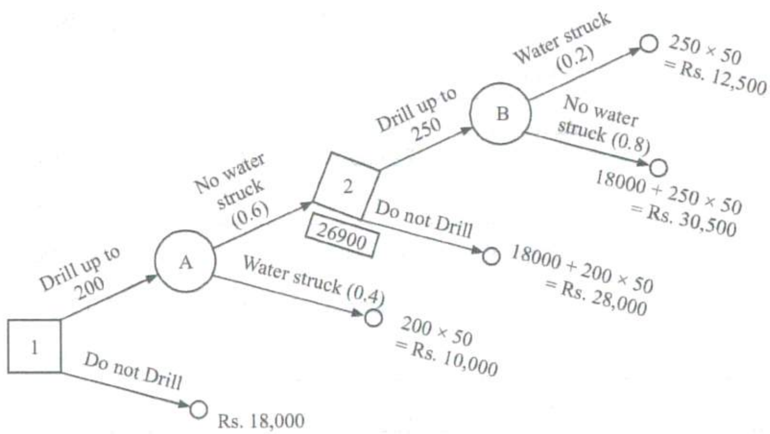
Example 3.19: A finance manager has to decide whether or not to drill a well on his farm. In his village, only 40 per cent of the wells drilled were successful at 200 feet of depth. Some of the farmers who did not get water at 200 feet , drilled further upto 250 feet but only 20 per cent struck water. Cost of drilling is ₹ 50 per foot. The finance manager estimated that he would pay ₹ 18,000 during a five year period on the present value terms, if he continues to buy water from the neighbour rather than go for the well, which would have a life of five years. He has three decisions to choose from: (*i*) Should he drill up to 200 feet, (*ii*) If no water is found at 200 feet, should he drill up to 250 feet? and (*iii*) Should he continue to buy water from his neighbour?

NOTES



Solution:

1. Draw the decisions tree diagram for the given problem.



2. Compute the associated cost of each outcome and enter these values on the decision tree.

$$\text{EMV of node } B = 0.2 \times 12500 + 0.8 \times 30500$$
$$= ₹\, 26{,}900.$$
$$\text{EMV of node } 2 = \text{Min}\,[26900, 28000]$$
$$= ₹\, 26{,}900.$$
$$\text{EMV of node } A = 0.4 \times 10000 + 0.6 \times 26900$$
$$= ₹\, 20{,}140.$$
$$\text{EMV of node } 1 = \text{Min}\,[20140, 18000]$$
$$= ₹\, 18{,}000.$$

Conclusion: Thus the optimal (least) course of action for the manager is not to drill the well and pay ₹ 18,000 for water to his neighbour for five years.

Advantages and Limitations of Decisions Tree Approach

Advantages

1. It structures the decision process and helps decision-making in a systematic and sequential order.
2. It helps the decision-maker to examine all possible outcomes, whether desirable or not.
3. It communicates the decision-making process to others in an easy and clear manner about the future.
4. It is mainly useful in situations where the initial decision and its outcome affects the subsequent decisions.
5. It displays the logical relationship between the parts of a complex decision and identifies the time sequence is which various action and subsequent events would occur.

Limitations

1. Decision tree diagrams become more complicated as the number of decision alternatives and variables increase.
2. It becomes highly complicated when interdependent alternatives and dependent variables are present in the problem.
3. It analyzes the expected values and gives an average solution only.
4. Often there is inconsistency in assigning probabilities for different events.

3.8.1 Rolling Back Techniques

Decision Trees for Sequential Decisions

Decision tree approach is a technique for making decision(s) especially in more complex risk situations. More complex decision problems can be solved conveniently using decision tree technique.

A decision tree (called tree as it looks like a tree) is a decision flow diagram that incorporates branches leading to alternatives one can select (the decision branches, often shown as dotted lines) among the usual branches leading to events that depend on probabilities (the probability branches, often shown as unbroken lines). In other words, a decision tree is a graphic way of showing the sequences of action-event combinations that are available to a decision-maker. Each sequence is shown by a distinct path through the tree. In the decision tree approach, we generally use the expectation principle, i.e., we choose the alternative that minimizes expected profit or the alternative that minimizes expected cost. The objective of using a decision tree is to decide which of the available alternative should in fact be selected.

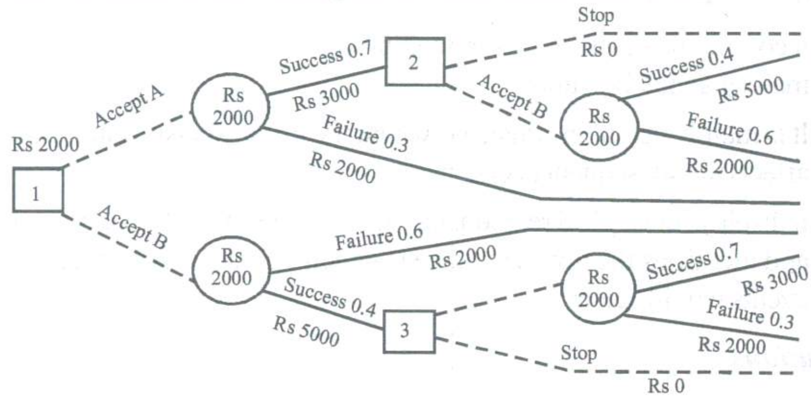
The preparation of a decision tree can be better understood by a decision tree problem. Let us take some decision problems for this purpose.

Example 3.20: A businessman has to select two independent investments A and B available to him but he lacks to undertake the capital of both of them simultaneously.

NOTES

He can choose *A* first and then stop, or if *A* is successful then take *B*, or vice versa. The probability of success on *A* is 0.7, while for *B* it is 0.4. Both investments require an initial capital outlay of ₹ 2000 and both return nothing if the venture is unsuccessful. Successful completion of *A* will return ₹ 3000 (over cost), successful completion of *B* will return ₹ 5000 (over cost). Draw the decision tree and determine the best strategy.

Solution: The decision tree of the problem can be drawn as follows:



The squares marked 1, 2 and 3 are the decision points and the circles represent event nodes. The expected payoff values have been shown in the circles. The optimal decision is to accept *A* first and if successful, then accept *B* as the expected payoff of this decision happens to be the maximum possible.

Explanation

To construct the above decision tree, the technique consists of the forward pass, the backward pass and finally the reading of the tree.

The Forward Pass

In this method, we first draw a square (the first decision point) on the left hand side of the tree. Alternatives available at this point (in our case *A* and *B*) are shown in dotted lines (the decision branches) leading to circles (the event nodes) from where the unbroken lines (the probability branches) are drawn (representing success and failures of *A* and also of *B* in our case). The probabilities and the related pay-offs are indicated on the probability branches. The probability branches representing success of *A* and *B* reach decision points 2 and 3 respectively. At the decision point 2, we may either stop or accept B and at decision point 3 we may either stop or accept *A*. The facts are shown as dotted lines and they lead us to new chance nodes from where again probability branches are drawn as required to complete the tree for the problem.

The Backward Pass

The next step in a decision tree analysis is to make the backward pass. To do this, we write against every final event on the tree, the value of that event. In a backward pass, the following rules are to be complied with:

(*i*) If the branches are probability branches, write the expectation in the relevant junction (or the node) of the branches.

(*ii*) If the branches are decision branches, select the branch with the highest expectation as the branch to be chosen and write this expectation alongside the junction of the branches (or the square). The selection may be indicated by ticking the decision branch.

By observing these rules, the relevant figures have been written in the circles and alongside the squares in the diagram of the tree drawn for the given problem.

NOTES

Check Your Progress

8. What is the aim of DDS?
9. What is a decision tree?

3.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The decision can be classifed as:
   (*i*) Technical decision
   (*ii*) Strategic decision
2. (*i*) Decision-making under certainty.
   (*ii*) Decision-making under uncertainty.
   (*iii*) Decision-making under risk.
   (*iv*) Decision-making under conflict.
3. The conditional value of each event in the payoff table is multiplied by its probability and the product is summed up. The resulting number is the EMV for the act.
4. The difference between the greater payoff and the actual payoff is known as opportunity loss.
5. The expected value with perfect information is the average (expected) return in the long run if we have perfect information before a decision is to be made.
6. If the decision-maker knows with certainty the consequences of every alternative or decision choice then the problem is a decision problem under certainty.
7. The following methods are available before the decision-maker in situations of uncertainty.
   (*i*) Maximax criterion
   (*ii*) Minimax criterion
   (*iii*) Maximin criterion
   (*iv*) Laplace criterion (criterion of equally likelihood)
   (*v*) Hurwicz alpha criterion (criterion of realism)

NOTES

8. Decision support systems/models (DSS) aim to provide tools that not only help such analyses, but enable the decision-maker to experiment and simulate the effects of different policies and selection strategies. The specific context of this model is administrative decision-making using large educational databases.
9. A decision tree (called tree as it looks like a tree) is a decision flow diagram that incorporates branches leading to alternatives one can select (the decision branches, often shown as dotted lines) among the usual branches leading to events that depend on probabilities (the probability branches, often shown as unbroken lines).

3.10 SUMMARY

- Decision-making is an everyday process in life. Right decisions will have salutary effect and the wrong one may prove to be disastrous.
- Decisions may be classified into two categories, tactical and strategic. Tactical decisions are those which affect the business in the short run. Strategic decisions are those which have far reaching effect on the course of business.
- The decision-maker is charged with the responsibility of making the decision. That is he has to select one from a set of possible courses of action.
- The acts are the alternative courses of action of strategies that are available to the decision-maker. Events are the occurrences which affect the achievement of the objectives. They are also called states of nature or outcomes.
- A payoff table represents the economics of a problem, i.e., revenue and costs associated with any action with a particular outcome. It is an ordered statement of profit or costs resulting under the given situation. The payoff can be interpreted as the outcome in quantitative form if the decision-maker adopts a particular strategy under a particular state of nature.
- An opportunity loss is the loss incurred because of failure to take the best possible action. Opportunity losses are calculated separately for each state of nature that might occur.
- In any decision problem, the decision-maker has to choose from the available alternative courses of action the one that yields the best result. If the consequences of each choice are known with certainty, the decision-maker can easily make decisions.
- In decision-making under certainty, the decision-maker knows with certainty the consequences of every alternative or decision choice. The decision-maker presumes that only one state of nature is relevant for his purpose.
- The term 'maximax' is an abbreviation of the phrase maximum of the maximums and an adventurous and aggressive decision-maker may choose to take the action that would result in the maximum payoff possible.

- Minimax is just the opposite of maximax. The losses are the costs to be incurred or the damages to be suffered for each of the alternative actions and states of nature. The minimax rule minimizes the maximum possible loss for each course of action. The term 'minimax' is an abbreviation of the phrase minimum of the maximum.
- The maximin criterion of decision-making stands for choice between alternative courses of action assuming pessimistic view of nature. Taking each act in turn, we note the worst possible results in terms of payoff and select the act which maximizes the minimum payoff.
- Laplace criterion is used when the decision-maker has no information about the probability of occurrence of various events. The decision-maker makes a simple assumption that each probability is equally likely. The expected payoff is worked out on the basis of these probabilities. The act having maximum expected payoff is selected.
- Hurwicz alpha criterion is a combination of maximin criterion and maximax criterion. In this method, the decision-maker's degree of optimism is represented by α, the coefficient of optimism. α varies between 0 and 1. When $\alpha = 0$, there is total pessimism and when $\alpha = 1$, there is total optimism.
- While decision-making under risk the decision-maker has to face several states of nature. The objective is to optimize the expected profit or to minimize the opportunity loss. For decision problems under risk, the most popular methods used are EMV (expected monetary value) criterion, EOL (expected opportunity loss), criterion or EVPI (expected value of perfect information).
- The conditional value of each event in the payoff table is multiplied by its probability and the product is summed up. The resulting number is the Expected Monetary Value (EMV) for the act. The decision-maker then selects from the available alternative actions, the action that leads to the maximum expected gain (that is the action with highest EMV).
- The difference between the greater payoff and the actual payoff is known as opportunity loss. Under this criterion, the strategy which has minimum Expected Opportunity Loss (EOL) is chosen. The calculation of EOL is similar to that of EMV.
- The Expected Value of Perfect Information (EVPI) is the average (expected) return in the long run, if we have perfect information before a decision is to be made. In order to calculate EVPI, we choose the best alternative with the probability of their state of nature. The EVPI is the expected outcome with perfect information minus the outcome with maximum EMV.

  EVPI = Expected value with perfect information – Maximum EMV
- The minimax regret criterion focuses on avoiding regrets that may result from making a non-optimal decision. It is an approach to make decision under uncertainty in which the opportunity cost which is termed as regret, associated with each possible course of action is measured and selects the activity that minimizes the maximum regret or loss is selected by the decision-

NOTES

maker. Regret is the difference between the best and worst possible payoff for each option.

- Deterministic model is related to deterministic situation. Deterministic decision pay-offs are the simplest possible payoffs. Probabilistic model or what is known as the stochastic decision model is related to risk situation. Perfect knowledge means that the decision-maker knows demand when he orders the goods. Competitive decision model is related to the situation of uncertainty.
- Decision tree approach is a technique for making decision(s) especially in more complex risk situations. It is a graphical representation of various decision alternatives, states of nature, probabilities attached to the states of nature and the conditional benefits as well as losses.
- In constructing a tree diagram, two types of nodes are used. Decision node represented by a square and State of nature node (chance node) represented by a circle. Alternative courses of action (strategies) start from the decision node as main branches. At the end of each main branch, there is a state of nature node from which emerge chance events in the form of sub-branches.
- The general approach used in decision tree analysis is to work backwards through the tree from right to left computing the expected value of each chance node. Then select the particular branch having a decision node, which leads to the chance node with the maximum expected value.

3.11 KEY TERMS

- **Decision:** It refers to selection among two or more alternative courses of action to achieve an objective.
- **Decision-maker:** Decision-maker is the person on whom lies the responsibility of making a decision and in selecting one from a set of possible courses of action.
- **Act:** It refers to the alternative course of action of strategies available to the decision-maker.
- **Event:** It is the occurrence affecting the achievement of objectives. This is also known as an outcome. The events constitute a mutually exclusive and exhaustive set of outcomes, which describe the possible behaviour of the environment in which the decision is made.
- **Payoff table:** It is a table that represents the economics of a problem, i.e., revenue and costs associated with any action with a particular outcome. It is an ordered statement of profit or costs resulting under the given situation. It is an outcome in quantitative form adopted by the decision-maker.
- **Opportunity loss table:** It is the loss incurred because of failure to take the best possible action. It is calculated separately for each state of nature that might occur.

- **Maximax decision criterion:** The term 'maximax' is an abbreviation of the phrase maximum of the maximums. It is an action chosen by the decision-maker that would result in the maximum possible payoff.
- **Minimax decision criterion:** Minimax is just the opposite of maximax. Application of the minimax criteria requires a table of losses instead of gains. The losses are the costs to be incurred or the damages to be suffered for each of the alternative actions and states of nature. The minimax rule minimizes the maximum possible loss for each course of action. The term 'minimax' is an abbreviation of the phrase minimum of the maximum.
- **Maximin decision criterion:** The maximin criterion of decision-making stands for choice between alternative courses of action assuming a pessimistic view of nature. Taking each act in turn, we note the worst possible results in terms of payoff and select the act which maximizes the minimum payoff.
- **Laplace criterion:** In the absence of information about the probability of occurrence of various events, the decision-maker makes a simple assumption that each probability is equally likely. The expected payoff is worked out on the basis of these probabilities. This is known as Laplace criterion.
- **Expected monetary value (EMV):** The statistical expectation of gain, for each course of action when probabilities can be assigned to the various states of nature is known as expected monetary value.
- **Expected opportunity loss (EOL):** The difference that exists between greater payoff and actual payoff. Under this criterion the strategy which has minimum expected opportunity loss (EOL) is chosen.
- **Expected value of perfect information (EVPI):** The expected value with perfect information is the average (expected) return in the long run, if we have perfect information before a decision is to be made.

3.12 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is an event?
2. What is a payoff table?
3. What is Laplace criterion?
4. Define the Hurwicz alpha criterion.
5. What is EMV?
6. What is minimax regret criterion?
7. What are the limitations of decision models?
8. What are the advantages of decision models?
9. Write two advantages of a decision tree.

Long-Answer Questions

1. Distinguish between decision under risk and decision under uncertainty.
2. Describe some methods which are useful for decision-making under uncertainty.
3. Explain the terms (*i*) Expected monetary value (*ii*) Expected value of perfect information with the help of examples.
4. Discuss EMV and EOL criteria.
5. What are payoff table and regret table? Explain with the help of examples.
6. Explain decision trees for sequential decisions. Why are they used?
7. What is the significance of action space? Explain with the help of an example.
8. Explain (*i*) Maximax (*ii*) Minimax, and (*iii*) Maximin decision criteria.
9. From the payoff table shown below decide which is the optimal act.

	Strategies		
Events	A_1	A_2	A_3
S_1	20	40	60
S_2	15	–10	–15
S_3	35	25	–20
	$P(S_1) = 0.4$	$P(S_2) = 0.5$	$P(S_3) = 0.1$

EMV for $A_1 = 19$, for $A_2 = 13.5$ and for $A_3 = 13.5$.

10. A businessman is trying to decide whether to take one of two contracts or none of them and then simplifying the situation into two alternatives. How can he make a decision tree? Explain giving suitable examples.

3.13 FURTHER READING

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House.

Gupta, S. C. 2006. *Fundamentals of Statistics*. New Delhi: Himalaya Publishing House.

Gupta, S. P., 2005. *Statistical Methods*. New Delhi: Sultan Chand and Sons.

Hooda, R. P. 2002. *Statistics for Business and Economics*. New Delhi: Macmillan India.

Kothari, C. R., 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Monga, G. S. 2000. *Mathematics and Statistics for Economics*. New Delhi: Vikas Publishing House

Gupta, S.P. 2006. *Statistical Methods*. New Delhi: S. Chand & Co. Ltd.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods*, 23rd edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Gupta, S.C. and V.K. Kapoor. *Fundamentals of Mathematical Statistics*. New Delhi: Sultan Chand & Sons.

Sharma, A. 2009. *Operations Research.* New Delhi: Himalaya Publishing House.

Singh, J. K. 2009. *Business Mathematics*. New Delhi: Himalaya Publishing House.

Kalavathy, S. 2010. *Operations Research with C Programmes,* 3rd edition. New Delhi: Vikas Publishing House Pvt. Ltd.

Kothari, C. R. 1992. *An Introduction to Operational Research.* New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 4 LINEAR PROGRAMMING

NOTES

Structure

4.0 INTRODUCTION

In this unit, you will learn the use of linear programming in decision-making. For a manufacturing process, a production manager has to take decisions as to what quantities and which process or processes are to be used so that the cost is minimum and profit is maximum. Currently, this method is used in solving a wide range of practical business problems. The word 'linear' means that the relationships are represented by straight lines. The word 'programming' means following a method for taking decisions systematically.

You will learn the extensive use of Linear Programming (LP) in solving resource allocation problems, production planning and scheduling, transportation, sales and advertising, financial planning, portfolio analysis, corporate planning, etc. Linear programming has been successfully applied in agricultural and industrial applications.

You will also learn a few basic terms like linearity, process and its level, criterion function, constraints, feasible solutions, optimum solution, etc. The

NOTES

term linearity implies straight line or proportional relationships among the relevant variables. Process means the combination of one or more inputs to produce a particular output. Criterion function is an objective function which is to be either maximized or minimized. Constraints are limitations under which one has to plan and decide. There are restrictions imposed upon decision variables. Feasible solutions are all those possible solutions considering given constraints. An optimum solution is considered the best among feasible solutions.

You will also learn to formulate linear programming problems and put these in a matrix form. The objective function, the set of constraints and the non-negative constraint together form a linear programming problem. In this unit, you will also learn the methods of solving a Linear Programming Problem (LPP) with two decision variables using the graphical method. All linear programming problems may not have unique solutions. You may find some linear programming problems that have an infinite number of optimal solutions, unbounded solutions or even no solution.

Finally, in this unit, you will learn about the canonical or standard form of LPP and duality. In the standard form, irrespective of the objective function, namely maximize or minimize, all the constraints are expressed as equations. Moreover, the right hand side (RHS) of each constraint and all variables are non-negative. The simplex method and M method are the methods of solution by iterative procedure in a finite number of steps using matrix.

4.1 UNIT OBJECTIVES

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

- Explain linear programming
- Discuss the terms associated with a linear programming problem
- Anslyse a matrix of a linear programming problem
- Describe the applications and limitations of linear programming problems
- Solve a linear programming problem with two variables using the graphical method, simplex method and M method
- Describe linear programming problems in canonical form
- Explain the concept of duality and solve dual problems

4.2 LINEAR PROGRAMMING: MEANING AND USES

Decision-making has always been very important in the business and industrial world, particularly with regard to the problems concerning production of commodities. Which commodity/commodities to produce, in what quantities and by which process or processes, are the main questions before a production

NOTES

manager. English economist Alfred Marshall pointed out that the businessman always studies his production function and his input prices and substitutes one input for another till his costs become the minimum possible. All this sort of substitution, in the opinion of Marshall, is being done by businessman's trained instinct rather than with formal calculations. But now there does exist a method of formal calculations often termed as Linear Programming. This method was first formulated by a Russian mathematician L.V. Kantorovich, but it was developed later in 1947 by George B. Dantzig 'for the purpose of scheduling the complicated procurement activities of the United States Air Force'. Today, this method is being used in solving a wide range of practical business problems. The advent of electronic computers has further increased its applications to solve many other problems in industry. It is being considered as one of the most versatile management tools.

4.2.1 Meaning of Linear Programming

Linear Programming (LP) is a major innovation since World War II in the field of business decision-making, particularly under conditions of certainty. The word 'Linear' means that the relationships are represented by straight lines, i.e., the relationships are of the form $y = a + bx$ and the word 'Programming' means taking decisions systematically. Thus, LP is a decision-making technique under given constraints on the assumption that the relationships amongst the variables representing different phenomena happen to be linear. In fact, Dantzig originally called it 'programming of interdependent activities in a linear structure' but later shortened it to 'Linear Programming'. LP is generally used in solving maximization (sales or profit maximization) or minimization (cost minimization) problems subject to certain assumptions. Putting in a formal way, 'Linear Programming is the maximization (or minimization) of a linear function of variables subject to a constraint of linear inequalities.' Hence, LP is a mathematical technique designed to assist the organization in optimally allocating its available resources under conditions of certainty in problems of scheduling, product-mix and so on.

4.2.2 Fields Where Linear Programming can be Used

The problem for which LP provides a solution may be stated to maximize or minimize for some dependent variable which is a function of several independent variables when the independent variables are subject to various restrictions. The dependent variable is usually some economic objectives, such as profits, production, costs, work weeks, tonnage to be shipped, etc. More profits are generally preferred to less profits and lower costs are preferred to higher costs. Hence, it is appropriate to represent either maximization or minimization of the dependent variable as one of the firm's objective. LP is usually concerned with such objectives under given constraints with linearity assumptions. In fact, it is powerful to take in its stride a wide range of business applications. The applications of LP are numerous and are increasing every day. LP is extensively used in solving resource allocation problems. Production planning and scheduling, transportation, sales and advertising, financial planning, portfolio analysis, corporate planning, etc., are some of its most fertile application areas. More specifically, LP has been successfully applied in the following fields:

NOTES

(*i*) *Agricultural applications:* LP can be applied in farm management problems as it relates to the allocation of resources such as acreage, labour, water supply or working capital in such a way that is maximizes net revenue.

(*ii*) *Contract awards:* Evaluation of tenders by recourse to LP guarantees that the awards are made in the cheapest way.

(*iii*) *Industrial applications:* Applications of LP in business and industry are of most diverse kind. Transportation problems concerning cost minimization can be solved by this technique. The technique can also be adopted in solving the problems of production (product-mix) and inventory control.

Thus, LP is the most widely used technique of decision-making in business and industry in modern times in various fields as stated above.

4.3 CONCEPTS, NOTATIONS AND GENERAL FORM OF LINEAR PROGRAMMING MODEL

The following are the concepts, notations and forms used in a linear programming model:

4.3.1 Basic Concepts and Notations

There are certain basic concepts and notations to be first understood for easy adoption of the LP technique. A brief mention of such concepts is as follows:

1. *Linearity.* The term linearity implies straight line or proportional relationships among the relevant variables. Linearity in economic theory is known as constant returns which means that if the amount of input doubles, the corresponding output and profit are also doubled. Linearity assumption, thus, implies that if two machines and two workers can produce twice as much as one machine and one worker; four machines and four workers twice as much as two machines and two workers and so on.

2. *Process and its level.* Process means the combination of particular inputs to produce a particular output. In a process, factors of production are used in fixed ratios, of course, depending upon technology and as such no substitution is possible with a process. There may be many processes open to a firm for producing a commodity and one process can be substituted for another. There is, thus, no interference of one process with another when two or more processes are used simultaneously. If a product can be produced in two different ways, then there are two different processes (or activities or decision variables) for the purpose of a linear programme.

NOTES

3. *Criterion function.* Criterion function is also known as objective function which states the determinants of the quantity either to be maximized or to be minimized. For example, revenue or profit is such a function when it is to be maximized or cost is such a function when the problem is to minimize it. An objective function should include all the possible activities with the revenue (profit) or cost coefficients per unit of production or acquisition. The goal may be either to maximize this function or to minimize this function. In symbolic form, let ZX denote the value of the objective function at the X level of the activities included in it. This is the total sum of individual activities produced at a specified level. The activities are denoted as $j = 1, 2, \ldots, n$. The revenue or cost coefficient of the jth activity is represented by C_j. Thus, $2X_1$, implies that X units of activity $j = 1$ yields a profit (or loss) of $C_1 = 2$.

4. *Constraints or inequalities.* These are the limitations under which one has to plan and decide, i.e., restrictions imposed upon decision variables. For example, a certain machine requires one worker to be operated upon; another machine requires at least four workers (i.e., > 4); there are at most 20 machine hours (i.e., < 20) available; the weight of the product should be say 10 lbs and so on, are all examples of constraints or why are known as inequalities. Inequalities like $X > C$ (reads X is greater than C or $X < C$ (reads X is less than C) are termed as strict inequalities. The constraints may be in form of weak inequalities like $X \leq C$ (reads X is less than or equal to C) or $X \geq C$ (reads C is greater than or equal to C). Constraints may be in the form of strict equalities like $X = C$ (reads X is equal to C).

   Let b_i denote the quantity b of resource i available for use in various production processes. The coefficient attached to resource i is the quantity of resource i required for the production of one unit of product j.

5. *Feasible solutions.* Feasible solutions are all those possible solutions which can be worked upon under given constraints. The region comprising of all feasible solutions is referred as *Feasible Region.*

6. *Optimum solution.* Optimum solution is the best of the feasible solutions.

4.3.2 General Form of the Linear Programming Model

Linear Programming problem mathematically can be stated as under:

Choose the quantities,

$$X_j \geq 0 \quad (j = 1, \ldots, n) \qquad \ldots(4.1)$$

This is also known as the non-negativity condition and in simple terms means that no X can be negative.

To maximize,

$$Z = \sum_{j=1}^{n} C_j X_j \qquad \ldots(4.2)$$

Subject to the constraints,

$$\sum_{j=1}^{n} a_{ij} X_j \le b_i \qquad (i = 1,...,m) \qquad ...(4.3)$$

NOTES

The above is the usual structure of a linear programming model in the simplest possible form. This model can be interpreted as a profit maximization situation where n production activities are pursued at level X_j which have to be decided upon, subject to a limited amount of m resources being available. Each unit of the jth activity yields a return C and uses an amount a_{ij} of the ith resource. Z denotes the optimal value of the objective function for a given system.

Assumptions or the conditions to be fulfilled underlying the LP model

LP model is based on the assumptions of proportionality, additivity, certainty, continuity and finite choices.

Proportionality is assumed in the objective function and the constraint inequalities. In economic terminology this means that there are constant returns to scale, i.e., if one unit of a product contributes ₹ 5 toward profit, then 2 units will contribute ₹ 10, 4 units ₹ 20 and so on.

Certainty assumption means the prior knowledge of all the coefficients in the objective function, the coefficients of the constraints and the resource values. LP model operates only under conditions of certainty.

Additivity assumption means that the total of all the activities is given by the sum total of each activity conducted separately. For example, the total profit in the objective function is equal to the sum of the profit contributed by each of the products separately.

Continuity assumption means that the decision variables are contiunous. Accordingly the combinations of output with fractional values, in case of product-mix problems, are possible and obtained frequently.

Finite choices assumption implies that finite number of choices are available to a decision-maker and the decision variables do not assume negative values.

4.4 APPLICATIONS AND LIMITATIONS OF LINEAR PROGRAMMING PROBLEMS

The applications of linear programming problems are based on linear programming matrix coefficients and data transmission prior to solving the simplex algorithm. The problem can be formulated from the problem statement using linear programming techniques. The following are the objectives of linear programming:

- Identify the objective of the linear programming problem, i.e., which quantity is to be optimized. For example, maximize the profit.
- Identify the decision variables and constraints used in linear programming, for example, production quantities and production limitations are taken as decision variables and constraints.

- Identify the objective functions and constraints in terms of decision variables using information from the problem statement to determine the proper coefficients.
- Add implicit constraints, such as non-negative restrictions.
- Arrange the system of equations in a consistent form and place all the variables on the left side of the equations.

NOTES

Applications of Linear Programming

Linear programming problems are associated with the efficient use of allocation of limited resources to meet desired objectives. A solution required to solve the linear programming problem is termed as optimal solution. The linear programming problems contain a very special subclass and depend on mathematical model or description. It is evaluated using relationships and are termed as straight-line or linear. The following are the applications of linear programming:

- Transportation problem
- Diet problem
- Matrix games
- Portfolio optimization
- Crew scheduling

Linear programming problem may be solved using a simplified version of the simplex technique called transportation method. Because of its major application in solving problems involving several product sources and several destinations of products, this type of problem is frequently called the transportation problem. It gets its name from its application to problems involving transporting products from several sources to several destinations. The formation is used to represent more general assignment and scheduling problems as well as transportation and distribution problems. The two common objectives of such problems are:

- To minimize the cost of shipping m units to n destinations.
- To maximize the profit of shipping m units to n destinations.

The goal of the diet problem is to find the cheapest combination of foods that will satisfy all the daily nutritional requirements of a person. The problem is formulated as a linear program where the objective is to minimize cost and meet constraints which require that nutritional needs be satisfied. The constraints are used to regulate the number of calories and amounts of vitamins, minerals, fats, sodium and cholesterol in the diet.

Game method is used to turn a matrix game into a linear programming problem. It is based on the Min-Max theorem which suggests that each player determines the choice of strategies on the basis of a probability distribution over the player's list of strategies.

The portfolio optimization template calculates the optimal capital of investments that gives the highest return for the least risk. The unique design of the portfolio optimization technique helps in financial investments or business portfolios.

The optimization analysis is applied to a portfolio of businesses to represent a desired and beneficial framework for driving capital allocation, investment and divestment decisions.

NOTES

Crew scheduling is an important application of linear programming problem. It helps if any airline has a problem related to a large potential crew schedules variables. Crew scheduling models are a key to airline competitive cost advantage these days because crew costs are the second largest flying cost after fuel costs.

Limitations of Linear Programming Problems

Linear programming is applicable if constraints and objective functions are linear, but there are some limitations of this technique which are as follows:

- All the uncertain factors, such as weather conditions, growth rate of industry, etc. are not taken into consideration.
- Integer values are not taken as the solution, e.g., a value is required for fraction and the nearest integer is not taken for the optimal solution.
- Linear programming technique gives those practical-valued answers that are really not desirable with respect to linear programming problem.
- It deals with one single objective in real life problem which is more limited and the problems come with multi-objective.
- In linear programming, coefficients and parameters are assumed as constants but in realty they do not take place.
- Blending is a frequently encountered problem in linear programming. For example, if different commodities are purchased which have different characteristics and costs, then the problem helps to decide how much of each commodity would be purchased and blended within specified bound so that the total purchase cost is minimized.

Check Your Progress

1. What is linear programming?
2. What is meant by criterion function in linear programming?
3. Mention two areas where linear programming finds application.
4. What are the constraints in linear programming?

4.5 FORMULATION OF LINEAR PROGRAMMING PROBLEM

A linear programming is a mathematical method for determining method to achieve the best outcome, i.e., maximum profit at lowest cost.

4.5.1 Graphic Solution

NOTES

The procedure for mathematical formulation of an LPP consists of the following steps:

Step 1: The decision variables of the problem are noted.

Step 2: The objective function to be optimized (maximized or minimized) as a linear function of the decision variables is formulated.

Step 3: The other conditions of the problem such as resource limitation, market constraints, interrelations between variables, etc., are formulated as linear inequations or equations in terms of the decision variables.

Step 4: The non-negativity constraint from the considerations is added so that the negative values of the decision variables do not have any valid physical interpretation.

The objective function, the set of constraints and the non-negative constraint together form a linear programming problem.

4.5.2 General Formulation of Linear Programming Problem

The general formulation of the LPP can be stated as follows:

In order to find the values of n decision variables $X_1 X_2 \ldots X_n$ to maximize or minimize the objective function.

$$Z = C_1X_1 + C_2X_2 + \ldots + C_nX_n \qquad \ldots (4.4)$$

$$\left.\begin{array}{l} a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n (\leq, =, \geq) b_1 \\ a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n (\leq, =, \geq) b_2 \\ \vdots \\ a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{in}X_n (\leq, =, \geq) b_i \\ \vdots \\ a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n (\leq = \geq) b_m \end{array}\right\} \qquad \ldots (4.5)$$

Here, the constraints can be inequality $\leq$ or $\geq$ or even in the form an equation (=) and finally satisfy the non-negative restrictions:

$$X_1 \geq 0,\ X_2 \geq 0 \ldots X_n \geq 0 \qquad \ldots (4.6)$$

4.5.3 Matrix Form of Linear Programming Problem

The LPP can be expressed in the matrix form as follows:

Maximize or minimize $Z = CX \rightarrow$ Objective function

Subject to $AX (\leq, =, \geq) B \rightarrow$ Constant equation

$B > 0, X \geq 0 \rightarrow$ Non-negativity restrictions

Where, $X = (X_1, X_2 \cdots X_n)$

$C = (C_1, C_2 \cdots C_n)$

NOTES

$$B = \begin{pmatrix} b_1 \\ b_2 \\ b_m \end{pmatrix} \quad A = \begin{pmatrix} a_{11}a_{12}\cdots a_{1n} \\ a_{21}a_{22}\cdots a_{2n} \\ : \\ a_{m1}a_{m2}\cdots a_{mn} \end{pmatrix}$$

Example 4.1: A manufacturer produces two types of models M_1 and M_2. Each model of the type M_1 requires 4 hours of grinding and 2 hours of polishing; whereas each model of the type M_2 requires 2 hours of grinding and 5 hours of polishing. The manufacturers have 2 grinders and 3 polishers. Each grinder works 40 hours a week and each polisher works for 60 hours a week. The profit on M_1 model is ₹ 3.00 and on model M_2 is ₹ 4.00. Whatever is produced in a week is sold in the market. How should the manufacturer allocate his production capacity to the two types of models, so that he may make the maximum profit in a week?

Solution:

Decision variables: Let X_1 and X_2 be the number of units of M_1 and M_2.

Objective function: Since the profit on both the models are given, we have to maximize the profit, viz.,

Max $Z = 3X_1 + 4X_2$

Constraints: There are two constraints: one for grinding and the other for polishing.

The number of hours available on each grinder for one week is 40 hours. There are 2 grinders. Hence, the manufacturer does not have more than $2 \times 40 =$ 80 hours for grinding. M_1 requires 4 hours of grinding and M_2 requires 2 hours of grinding.

The grinding constraint is given by,

$$4X_1 + 2X_2 \leq 80$$

Since there are 3 polishers, the available time for polishing in a week is given by $3 \times 60 = 180$. M_1 requires 2 hours of polishing and M_2 requires 5 hours of polishing. Hence, we have $2X_1 + 5X_2 \leq 180$

Thus, we have,

Max $Z = 3X_1 + 4X_2$

Subject to, $4X_1 + 2X_2 \leq 80$

$$2X_1 + 5X_2 \leq 180$$

$$X_1, X_2 \geq 0$$

Example 4.2: A firm manufacturers three products A, B and C. The profits are ₹ 3, ₹ 2 and ₹ 4 respectively. The firm has two machines and the following is the required processing time in minutes for each machine on each product.

		Product		
		A	*B*	*C*
Machines	*C*	4	3	5
	D	3	2	4

NOTES

Machine *C* and *D* have 2000 and 2500 machine minutes respectively. The firm must manufacture 100 units of *A*, 200 units of *B* and 50 units of *C*, but not more than 150 units of *A*. Set up an LP problem to maximize the profit.

Solution: Let X_1, X_2, X_3 be the number of units of the product *A, B, C* respectively.

Since the profits are ₹ 3, ₹ 2 and ₹ 4 respectively, the total profit gained by the firm after selling these three products is given by,

$$Z = 3X_1 + 2X_2 + 4X_3$$

The total number of minutes required in producing these three products at machine *C* is given by $4X_1 + 3X_2 + 5X_3$ and at machine *D* is given by,

$3X_1 + 2X_2 + 4X_3$.

The restrictions on the machine *C* and *D* are given by 2000 minutes and 2500 minutes.

$$4X_1 + 3X_2 + 5X_3 \leq 2000$$
$$3X_1 + 2X_2 + 4X_3 \leq 2500$$

Also, since the firm manufactures 100 units of *A*, 200 units of *B* and 50 units of *C*, but not more than 150 units of *A*, the further restriction becomes,

$$100 \leq X_1 \leq 150$$
$$200 \leq X_2 \geq 0$$
$$50 \leq X_3 \geq 0$$

Hence, the allocation problem of the firm can be finally put in the following form:

Find the value of X_1, X_2, X_3 so as to maximize,

$Z = 3X_1 + 2X_2 + 4X_3$

Subject to the constraints,

$$4X_1 + 3X_2 + 5X_3 \leq 2000$$
$$3X_1 + 2X_2 + 4X_3 \leq 2500$$

$$100 \leq X_1 \leq 150, 200 \leq X_2 \geq 0, 50 \leq X_3 \geq 0$$

NOTES

Check Your Progress

5. What do you understand by solution in the context of linear programming problem?
6. What is a 'basic solution' of an LPP?
7. What is meant by basic and non-basic variables?
8. What do you understand by basic feasible solution?

4.6 SOLUTION OF LINEAR PROGRAMMING PROBLEM: GRAPHICAL SOLUTION AND SIMPLEX METHOD

The linear programming problems can be solved as follows using the graphical solution and simplex method:

4.6.1 Graphical Solution

Simple linear programming problem with two decision variables can be easily solved by graphical method.

Procedure for solving LPP by graphical method

The steps involved in the graphical method are as follows:

Step 1: Consider each inequality constraint as an equation.

Step 2: Plot each equation on the graph as each will geometrically represent a straight line.

Step 3: Mark the region. If the inequality constraint corresponding to that line is $\leq$, then the region below the line lying in the first quadrant (due to non-negativity of variables) is shaded. For the inequality constraint $\geq$ sign, the region above the line in the first quadrant is shaded. The points lying in the common region will satisfy all the constraints simultaneously. The common region thus obtained is called the feasible region.

Step 4: Allocate an arbitrary value, say zero, for the objective function.

Step 5: Draw the straight line to represent the objective function with the arbitrary value (i.e., a straight line through the origin).

Step 6: Stretch the objective function line till the extreme points of the feasible region. In the maximization case, this line will stop farthest from the origin and passes through at least one corner of the feasible region. In the minimization case this line will stop nearest to the origin and passes through at least one corner of the feasible region.

Step 7: Find the coordinates of the extreme points selected in Step 6 and find the maximum or minimum value of Z.

Note: As the optimal values occur at the corner points of the feasible region, it is enough to calculate the value of the objective function of the corner points of the feasible region and select the one which gives the optimal solution, i.e., in the case of maximization problem, optimal point corresponds to the corner point at which the objective function has a maximum value and in the case of minimization, the corner point which gives the objective function the minimum value is the optimal solution.

NOTES

Example 4.3: Solve the following LPP by graphical method.

Minimize, $Z = 20X_1 + 10X_2$

Subject to, $X_1 + 2X_2 \leq 40$

$3X_1 + X_2 \geq 30$

$4X_1 + 3X_2 \geq 60$

$X_1, X_2 \geq 0$

Solution: Replace all the inequalities of the constraints by equation,

$X_1 + 2X_2 = 40$ If $X_1 = 0 \Rightarrow X_2 = 20$

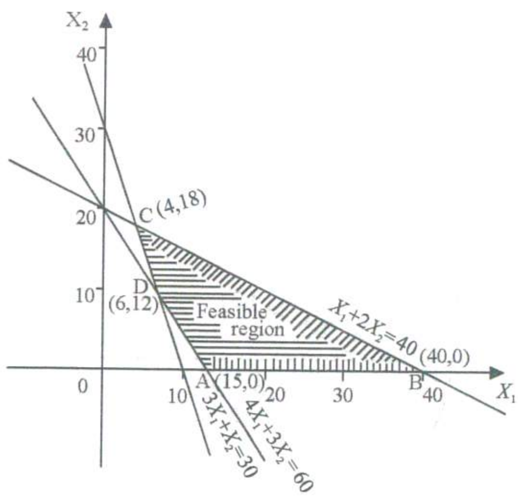
If $X_2 = 0 \Rightarrow X_1 = 40$

$\therefore X_1 + 2X_2 = 40$ passes through (0, 20) (40, 0)

$3X_1 + X_2 = 30$ passes through (0, 30) (10, 0)

$4X_1 + 3X_2 = 60$ passes through (0, 20) (15, 0)

Plot each equation on the graph.



The feasible region is *ABCD*.

C and *D* are points of intersection of lines.

$X_1 + 2X_2 = 40,\ 3X_1 + X_2 = 30$

And, $4X_1 + 3X_2 = 60$

On solving, we get *C* (4, 18) and *D* (6, 12)

Corner Points	*Value of* $Z = 20X_1 + 10X_2$
A (15, 0)	300
B (40, 0)	800
C (4, 18)	260

$D\,(6, 12)$ 240 (Minimum value)

$\therefore$ The minimum value of Z occurs at $D\,(6, 12)$. Hence, the optimal solution is $X_1 = 6, X_2 = 12$.

NOTES

Example 4.4: Find the maximum value of $Z = 5X_1 + 7X_2$

Subject to the constraints,

$X_1 + X_2 \leq 4$

$3X_1 + 8X_2 \leq 24$

$10X_1 + 7X_2 \leq 35$

$X_1, X_2 > 0$

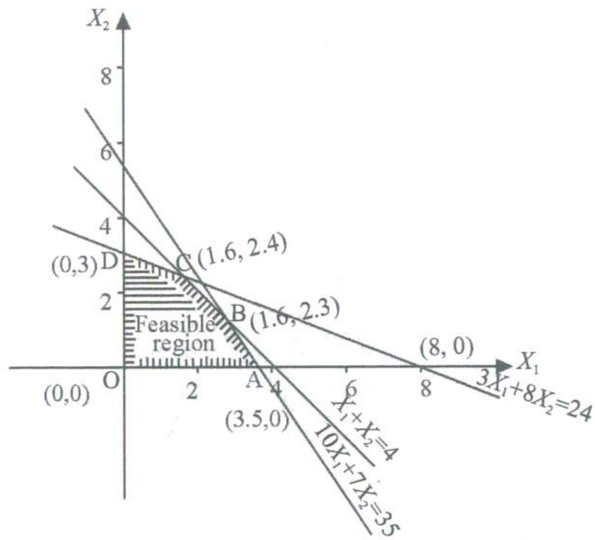
Solution: Replace all the inequalities of the constraints by forming equations.

$X_1 + X_2 = 4$ passes through (0, 4) (4, 0)

$3X_1 + 8X_2 = 24$ passes through (0, 3) (8, 0)

$10X_1 + 7X_2 = 35$ passes through (0, 5) (3.5, 0)

Plot these lines in the graph and mark the region below the line as the inequality of the constraint is $\leq$ and is also lying in the first quadrant.



The feasible region is $OABCD$.

B and C are points of instruction of lines,

$X_1 + X_2 = 4,\ 10X_1 + 7X_2 = 35$

And, $3X_1 + 8X_2 = 24$

On solving we get,

$B\,(1.6, 2.3)$

$C\,(1.6, 2.4)$

Corner Points	Value of $Z = 5X_1 + 7X_2$
O (0, 0)	0
A (3.5, 0)	17.5
B (1.6, 2.3)	25.1
C (1.6, 2.4)	24.8 (Maximum value)
D (0, 3)	21

NOTES

∴ The maximum value of Z occurs at C (1.6, 2.4) and the optimal solution is $X_1 = 1.6, X_2 = 2.4$.

Example 4.5: Solve the following LPP by graphical method.

$$\text{Maximize, } Z = 100X_1 + 40X_2$$

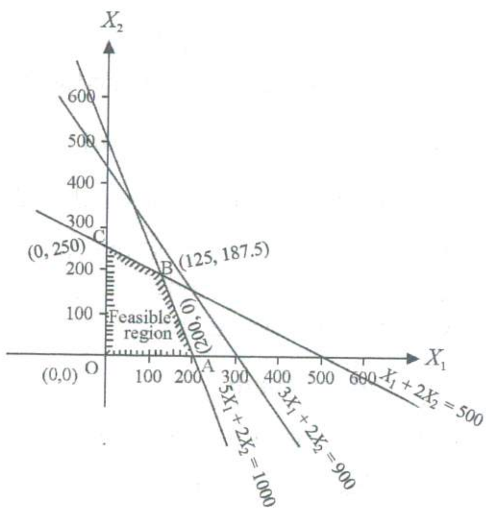
$$\text{Subject to, } 5X_1 + 2X_2 \leq 1000$$

$$3X_1 + 2X_2 \leq 900$$

$$X_1 + 2X_2 \leq 500$$

$$\text{and } X_1 + X_2 \geq 0$$

Solution:



The solution space is given by the feasible region $OABC$.

Corner Points	Value of $Z = 100X_1 + 40X_2$	
O (0, 0)	0	
A (200, 0)	20,000	
B (125, 187.5)	20,000	(Maximum value of Z)
C (0, 250)	10,000	

∴ The maximum value of Z occurs at two vertices A and B.

Since there are infinite number of points on the line, joining A and B gives the same maximum value of Z.

Thus, there are infinite number of optimal solutions for the LPP.

4.6.2 Some Important Definitions

NOTES

1. A set of values $X_1, X_2 \dots X_n$ which satisfies the constraints of the LPP is called its solution.
2. Any solution to a LPP which satisfies the non-negativity restrictions of the LPP is called its feasible solution.
3. Any feasible solution which optimizes (minimizes or maximizes) the objective function of the LPP is called its optimum solution.
4. Given a system of m linear equations with n variables ($m < n$), any solution which is obtained by solving m variables keeping the remaining $n - m$ variables zero is called a basic solution. Such m variables are called basic variables and the remaining variables are called non-basic variables.
5. A basic feasible solution is a basic solution which also satisfies all basic variables are non-negative.

Basic feasible solutions are of two types:

(*i*) *Non-degenerate:* A non-degenerate basic feasible solution is the basic feasible solution which has exactly m positive X_i ($i = 1, 2, \dots, m$), i.e., none of the basic variables is zero.

(*ii*) *Degenerate:* A basic feasible solution is said to be degenerate if one or more basic variables are zero.

6. If the value of the objective function Z can be increased or decreased indefinitely, such solutions are called unbounded solutions.

4.6.3 Canonical or Standard Forms of LPP

The general LPP can be put in either canonical or standard forms.

In the standard form, irrespective of the objective function, namely maximize or minimize, all the constraints are expressed as equations. Moreover, RHS of each constraint and all variables are non-negative.

Characteristics of the standard form

(*i*) The objective function is of maximization type.

(*ii*) All constraints are expressed as equations.

(*iii*) Right hand side of each constraint is non-negative.

(*iv*) All variables are non-negative.

In the canonical form, if the objective function is of maximization, all the constraints other than non-negativity conditions are '≤' type. If the objective function is of minimization, all the constraints other than non-negative conditions are '≥' type.

Characteristics of the canonical form

(*i*) The objective function is of maximization type.

(*ii*) All constraints are of '≤' type.

(*iii*) All variables X_i are non-negative.

Notes:

1. Minimization of a function Z is equivalent to maximization of the negative expression of this function, i.e., Min $Z = -\text{Max}\,(-Z)$.
2. An inequality in one direction can be converted into an inequality in the opposite direction by multiplying both sides by (-1).
3. Suppose we have the constraint equation,

$$a_{11}X_1 + a_{12}X_2 + \ldots\ldots + a_{1n}X_n = b_1$$

This equation can be replaced by two weak inequalities in opposite directions.

$$a_{11}X_1 + a_{12}X_2 + \ldots\ldots + a_m X_n \leq b_1$$

$$a_{11}X_1 + a_{12}X_2 + \ldots\ldots + a_{1n}X_n \geq b_1$$

4. If a variable is unrestricted in sign, then it can be expressed as a difference of two non-negative variables, i.e., X_1 is unrestricted in sign, then $X_i = X_i' - X_i''$, where X_i, X_i', X_i'' are ≥ 0.
5. In standard form, all the constraints are expressed in equation, which is possible by introducing some additional variables called slack variables and surplus variables so that a system of simultaneous linear equations is obtained. The necessary transformation will be made to ensure that $b_i \geq 0$.

NOTES

Definition

(*i*) If the constraints of a general LPP be,

$$\sum_{j=1}^{n} a_{ij}\, X_i \leq b_i \;(i = 1, 2, \ldots, m),$$

then the non-negative variables S_i, which are introduced to convert the inequalities ($\leq$) to the equalities $\sum_{j=1}^{n} a_{ij}\, X_i + S_i = b_i \;(i = 1, 2, \ldots, m)$, are called slack variables.

Slack variables are also defined as the non-negative variables which are added in the LHS of the constraint to convert the inequality '$\leq$' into an equation.

(*ii*) If the constraints of a general LPP be,

$$\sum_{j=1}^{n} a_{ij}\, X_j \geq b_i \;(i = 1, 2, \ldots, m),$$

then, the non-negative variables S_i which are introduced to convert the inequalities '$\geq$' to the equalities $\sum_{j=1}^{n} a_{ij}\, X_j - S_i = b_i \;(i = 1, 2, \ldots, m)$ are called surplus variables.

Surplus variables are defined as the non-negative variables which are removed from the LHS of the constraint to convert the inequality '$\geq$' into an equation.

4.6.4 Simplex Method

Simplex method is an iterative procedure for solving LPP in a finite number of steps. This method provides an algorithm which consists of moving from one vertex of the region of feasible solution to another in such a manner that the value of the objective function at the succeeding vertex is less or more as the case may be that at the previous vertex. This procedure is repeated and since the number of vertices is finite, the method leads to an optimal vertex in a finite number of steps or indicates the existence of unbounded solution.

NOTES

Definition

(*i*) Let X_B be a basic feasible solution to the LPP.

Max $Z = C_X$

Subject to $A_X = b$ and $X \geq 0$, such that it satisfies $X_B = B^{-1}b$,

Where B is the basic matrix formed by the column of basic variables.

The vector $C_B = (C_{B1}, C_{B2} \ldots C_{Bm})$, where C_{Bj} are components of C associated with the basic variables is called the cost vector associated with the basic feasible solution X_B.

(*ii*) Let X_B be a basic feasible solution to the LPP.

Max $Z = C_X$, where $A_X = b$ and $X \geq 0$.

Let C_B be the cost vector corresponding to X_B. For each column vector a_j in A_1, which is not a column vector of B, let

$$a_j = \sum_{i=1}^{m} a_{ij} b_j$$

Then the number $Z_j = \sum_{i=1}^{m} C_{Bi} a_{ij}$ is called the evaluation corresponding to a_j and the number $(Z_j - C_j)$ is called the net evaluation corresponding to j.

Simplex algorithm

For the solution of any LPP by simplex algorithm, the existence of an initial basic feasible solution is always assumed. The steps for the computation of an optimum solution are as follows:

Step 1: Check whether the objective function of the given LPP is to be maximized or minimized. If it is to be minimized then we convert it into a problem of maximization by,

Min $Z = -$Max $(-Z)$

Step 2: Check whether all b_i $(i = 1, 2, \ldots, m)$ are positive. If any one of b_i is negative, then multiply the inequation of the constraint by -1 so as to get all b_i to be positive.

Step 3: Express the problem in the standard form by introducing slack/surplus variables to convert the inequality constraints into equations.

Step 4: Obtain an initial basic feasible solution to the problem in the form $X_B = B^{-1}b$ and put it in the first column of the simplex table. Form the initial simplex table shown as follows:

		C_j	C_1	C_2	C_3	..	..	0 0 0
C_B	S_B	X_B	X_1	X_2	X_3	X_4	..... X_n	S_1.. S_2 ..... S_m
C_{B1}	S_1	b_1	a_{11}	a_{12}	a_{13}	a_{14}	..... a_{1n}	1 0 0
C_{B2}	S_2	b_2	a_{21}	a_{22}	a_{23}	a_{24}	..... a_{2n}	1 0 0

NOTES

Step 5: Compute the net evaluations $Z_j - C_j$ by using the relation:

$Z_j - C_j = C_B\,(a_j - C_j)$

Examine the sign of $Z_j - C_j$:

(*i*) If all $Z_j - C_j \geq 0$, then the initial basic feasible solution X_B is an optimum basic feasible solution.

(*ii*) If at least one $Z_j - C_j > 0$, then proceed to the next step as the solution is not optimal.

Step 6: To find the entering variable, i.e., key column.

If there are more than one negative $Z_j - C_j$ choose the most negative of them. Let it be $Z_r - C_r$ for some $j = r$. This gives the entering variable X_r and is indicated by an arrow at the bottom of the *r*th column. If there are more than one variable having the same most negative $Z_j - C_j$, then any one of the variable can be selected arbitrarily as the entering variable.

(*i*) If all $X_{ir} \leq 0$ $(i = 1, 2, \ldots, m)$ then there is an unbounded solution to the given problem.

(*ii*) If at least one $X_{ir} > 0$ $(i = 1, 2, \ldots, m)$, then the corresponding vector X_r enters the basis.

Step 7: To find the leaving variable or key row:

Compute the ratio $(X_{Bi}/X_{kr}, X_{ir} > 0)$

If the minimum of these ratios be X_{Bi}/X_{kr}, then choose the variable X_k to leave the basis called the key row and the element at the intersection of the key row and the key column is called the key element.

Step 8: Form a new basis by dropping the leaving variable and introducing the entering variable along with the associated value under C_B column. The leaving element is converted to unity by dividing the key equation by the key element and all other elements in its column to zero by using the formula:

New element = Old element

$$-\left[\frac{\text{Product of elements in key row and key column}}{\text{Key element}}\right]$$

Step 9: Repeat the procedure of Step (5) until either an optimum solution is obtained or there is an indication of unbounded solution.

Example 4.6: Use simplex method to solve the following LPP.

Maximize, $Z = 3X_1 + 2X_2$

NOTES

Subject to, $X_1 + X_2 \le 4$

$X_1 - X_2 \le 2$

$X_1, X_2 \ge 0$

Solution: By introducing the slack variables S_1 and S_2 convert the problem into standard form.

Max, $Z = 3X_1 + 2X_2 + 0S_1 + 0S_2$

Subject to, $X_1 + X_2 + S_1 = 4$

$X_1 - X_2 + S_2 = 2$

$X_1, X_2, S_1, S_2 \ge 0$

$$\begin{bmatrix} X_1 & X_2 & S_1 & S_2 \\ 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

An initial basic feasible solution is given by,

$X_B = B^{-1}b,$

Where, $B = I_2, X_B = (S_1, S_2)$

i.e., $(S_1, S_2) = I_2\,(4, 2) = (4, 2)$

Initial simplex table

$Z_j = C_B\, a_j$

$$Z_1 - C_1 = C_B a_1 - C_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} (1\ \ 1) - 3 = -3$$

$$Z_2 - C_2 = C_B a_2 - C_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} (1\ \ 1) - 2 = -2$$

$$Z_3 - C_3 = C_B a_3 - C_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} (1\ \ 0) - 0 = -0$$

$$Z_4 - C_4 = C_B a_4 - C_4 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} (0\ \ 1) - 0 = -0$$

		C_j	3	2	0	0	
C_B	B	X_B	X_1	X_2	S_1	S_2	Min $\frac{X_B}{X_1}$
0	S_1	4	1	1	1	0	4/1 = 4
←0	S_2	2	(1)	–1	0	1	2/1 = 2
	Z_j	0	0	0	0	0	
	$Z_j - C_j$		–3↑	–2	0	0	

NOTES

Since, there are some $Z_j - C_j = 0$, the current basic feasible solution is not optimum.

Since, $Z_1 - C_1 = -3$ is the most negative, the corresponding non-basic variable X_1 enters the basis.

The column corresponding to this X_1 is called the key column.

$$\text{Ratio} = \text{Min}\left\{\frac{X_{Bi}}{X_{ir}}, X_{ir} > 0\right\}$$

$$= \text{Min}\left\{\frac{4}{1}, \frac{2}{1}\right\}, \text{ which corresponds to } S_2$$

$\therefore$ The leaving variable is the basic variable S_2. This row is called the key row. Convert the leading element X_{21} to units and all other elements in its column n, i.e., (X_1) to zero by using the formula:

New element = Old element –

$$\left[\frac{\text{Product of elements in key row and key column}}{\text{Key element}}\right]$$

To apply this formula, first we find the ratio, namely

$$\frac{\text{The element to be zero}}{\text{Key element}} = \frac{1}{1} = 1$$

Apply this ratio for the number of elements that are converted in the key row. Multiply this ratio by key row element shown as follows:

1×2

1×1

1×-1

1×0

1×1

Now, subtract this element from the old element. The element to be converted into zero is called the old element row. Finally, we have

$4 - 1 \times 2 = 2$

$1 - 1 \times 1 = 0$

$1 - 1 \times -1 = 2$

$1 - 1 \times 0 = 1$

$0 - 1 \times 1 = -1$

$\therefore$ The improved basic feasible solution is given in the following simplex table.

First iteration

NOTES

		C_j	3	2	0	0	
C_B	B	X_B	X_1	X_2	S_1	S_2	Min $\frac{X_B}{X_2}$
←0	S_1	2	0	(2)	1	–1	2/2 = 1
3	X_1	2	1	–1	0	1	–
	Z_j	6	3	–3	0	0	
	$Z_j - C_j$		0	–5↑	0	0	

Since, $Z_2 - C_2$ is the most negative, X_2 enters the basis.

To find Min $\left(\frac{X_B}{X_{i2}}, X_{i2} > 0\right)$

$$\text{Min}\left(\frac{2}{2}\right)$$

This gives the outgoing variables. Convert the leaving element into one. This is done by dividing all the elements in the key row by 2. The remaining elements are converted to zero by using the following formula.

Here, $-\frac{1}{2}$ is the common ratio. Put this ratio 5 times and multiply each ratio by the key row element.

$-\frac{1}{2} \times 2$

$-\frac{1}{2} \times 0$

$-\frac{1}{2} \times 2$

$-1/2 \times 1$

$-1/2 \times -1$

Subtract this from the old element. All the row elements which are converted into zero are called the old elements.

$2 - \left(-\frac{1}{2} \times 2\right) = 3$

$1 - (-1/2 \times 0) = 1$

$-1 - (-1/2 \times 2) = 0$

$0 - (-1/2 \times 1) = 1/2$

$1 - (-1/2 \times -1) = 1/2$

NOTES

Second iteration

		C_j	3	2	0	0
C_B	B	X_B	X_1	X_2	S_1	S_2
2	X_2	1	0	1	1/2	–1/2
3	X_1	3	1	0	1/2	1/2
	Z_j	11	3	2	5/2	1/2
	Z_j–C_j		0	0	5/2	1/2

Since all $Z_j - C_j \geq 0$, the solution is optimum. The optimal solution is Max $Z = 11, X_1 = 3$, and $X_2 = 1$.

Example 4.7: Solve the following LPP using simplex method.

Maximize $Z=3X_1 + 2X_2 + 5X_3$

Subject to, $X_1 + 2X_2 + X_3 \leq 430$

$3X_1 + 2X_3 \leq 460$

$X_1 + 4X_2 \leq 420$

$X_1, X_2, X_3 \geq 0$

Solution: Rewrite the constraint into an equation by adding slack variables S_1, S_2 and S_3. The standard form of LPP becomes,

Maximize, $Z = 3X_1 + 2X_2 + 5X_3 + 0S_1 + 0S_2 + 0S_3$

Subject, to,$X_1 + 2X_2 + X_3 + S_1 = 430$

$3X_1 + 2X_3 + S_2 = 460$

$X_1 + 4X_2 + S_3 = 420$

$X_1, X_2, X_3, S_1, S_2, S_3 \geq 0$

The initial basic feasible solution is,

$S_1 = 430, S_2 = 460, S_3 = 420$ ($X_1 = X_2 = X_3 = 0$)

Initial table

		C_j	3	2	5	0	0	0	
C_B	B	X_B	X_1	X_2	X_3	S_1	S_2	S_3	Min $\frac{X_B}{X_3}$
0	S_1	430	1	2	1	1	0	0	430/1=430
←0	S_2	460	3	0	(2)	0	1	0	460/2=230
0	S_3	420	1	4	0	0	0	1	
	Z_j	0	0	0	0	0	0	0	
	$Z_j - C_j$		–3	–2	–5↑	0	0	0	

Since some of $Z_j - C_j \leq 0$, the current basic feasible solution is not optimum. Since $Z_3 - C_3 = -5$ is the most negative, the variable X_3 enters the basis. To find the variable leaving the basis, find

$$\text{Min}\left(\frac{X_B}{X_{i3}}, X_{i3} > 0\right) = \text{Min}\left(\frac{430}{1} = 430, \frac{460}{2} = 230\right)$$

NOTES

∴ The variable S_2 leaves the basis.

First iteration

		C_j	3	2	5	0	0	0	
C_B	B	X_B	X_1	X_2	X_3	S_1	S_2	S_3	Min $\frac{X_B}{X_2}$
←0	S_1	200	–1/2	(2)	0	1	1/2	0	200/2=100
5	X_3	230	3/2	0	1	0	1/2	0	
0	S_3	420	1	4	0	0	0	1	420/4=105
	Z_j	1150	15/2	0	0	0	5/2	0	
	$Z_j - C_j$		9/2	–2↑	0	0	5/2	0	

Since $Z_2 - C_2 = -2$ is negative, the current basic feasible solution is not optimum. Therefore, the variable X_2 enters the basis and the variable S_1 leaves the basis.

4.6.5 *M* Method

In simplex algorithm, the *M* Method is used to deal with the situation where an infeasible starting basic solution is given. The simplex method starts from one basic feasible solution (BFS) or the intense point of the feasible region of a linear programming problem (LPP) presented in tableau form and extends to another BFS for constantly raising or reducing the value of the objective task till optimality is reached. Sometimes the starting basic solution may be infeasible, then *M* method is used to find the starting basic feasible solution each time it exists.

Example 4.8: Find a starting basic feasible solution each time it exists for the following LPP where there is no starting identity matrix using *M* method.

Maximize, $X_0 = C^T X$

Subject to, $AX = b, X \geq 0$; Where $b > 0$.

Solution: To get a starting identity matrix, we add artificial variables $X_{a1}, X_{a2}, \ldots\ldots, X_{am}$. The consequent values for the artificial variables can be M for maximization problem (where M is adequately large). This constant M will check artificial variables that will arise with positive values in the final optimal solutions. Now the LPP becomes,

Max $Z = C^T X - M \,.\, 1^T X_a$

Subject to, $AX + I_m X_a = b$,

$X \geq 0$

Where $X_a = (X_{a1}, X_{a2}, \ldots\ldots, X_{am})^T$ and 1 is the vector of all ones. Here, $X = 0$ and $X_a = b$ is the feasible starting basic feasible solution. For solving $AX + I_m X_a = b$, which is a solution to $AX = b$ we have to drive and take $X_a = 0$.

NOTES

Example 4.9: Using the linear programming, solve the following LPP:

Maximize, $X_0 = X_1 + X_2$

Subject to, $2X_1 + X_2 \geq 4$

$X_1 + 2X_2 = 6$

$X_1, X_2 \geq 0$

Solution: Add surplus variable X_3 and artificial variables X_4 and X_5, and then rewrite the equation as given below:

$$2X_1 + X_2 - X_3 + X_4 = 4$$

$$X_1 + 2X_2 + X_5 = 6$$

$$X_0 - X_1 - X_2 + MX_4 + MX_5 = 0$$

The columns corresponding to X_4 and X_5 form an identity matrix. This can be represented in tableau form as,

	X_1	X_2	X_3	X_4	X_5	b
X_4	2	1	–1	1	0	4
X_5	1	2	0	0	1	6
X_0	–1	–1	0	M	M	0

In the above table the row X_0 has the reduced cost coefficient for basic variables X_4 and X_5 which are not zero. First eliminate these nonzero entries to have the initial tableau.

	X_1	X_2	X_3	X_4	X_5	b
X_4	2	1	–1	1	0	4
X_5	1	2	0	0	1	6
X_0	$-(1 + 3M)$	$-(1 + 3M)$	M	0	0	$-10M$

The artificial variable becomes non-basic and can be dropped in subsequent calculations. Now the tableau becomes:

	X_1	X_2	X_3	X_5	b
X_1	1	1/2	–1/2	0	2
X_5	0	3/2	1/2	1	4
X_0	0	$-(1 + 3M)/2$	$-(1 + M)/2$	0	$2 - 4M$

Eliminating artificial variables we get,

	X_1	X_2	X_3	b
X_1	1	0	–2/3	2/3
X_2	0	1	1/3	8/3
X_0	0	0	–1/3	10/3

NOTES

Now all the artificial variables are eliminated and $X = [2/3, 8/3, 0]^T$ is an initial basic feasible solution. Iterating again we get we following final optimal tableau:

	X_1	X_2	X_3	*b*
X_1	1	2	0	6
X_3	0	3	1	8
X_0	0	1	0	6

Hence, the optimal solution is $X = (6, 0, 8)^T$ with $X_0 = 6$.

Check Your Progress

9. When is an objective function minimized? When is it maximized?
10. What is meant by a feasible solution?
11. What is a feasible region?
12. What is an optimal solution?
13. What are non-degenerate and degenerate type basic feasible solutions?
14. What is the simplex method?
15. How is a leaving element converted to unity in a simplex algorithm?
16. What is the role of the slack variable?
17. When is M method used?

4.7 DUALITY

For every given linear programming problem, there is another intimately related linear programming problem referred to as its dual. The duality theorem states that 'for every maximization (or minimization) problem in linear programming, there is a unique similar problem of minimization (or maximization) involving the same date which describes the original problem'. The original problem is referred to as the 'primal'. The 'dual' of a dual problem is the primal. Thus the primal and dual problems are replicas of each other. Further, the maximum feasible value of the primal objective function equals to the minimum feasible value of the dual objective function. This means that the solutions of the primal and the dual problems are related which infact yields several advantages.

The transformation of a given primal problem into a dual problem involves the following considerations:

(1) If the objective of the primal is maximization, the objective of the dual is minimization.

(2) The primal has m-constraints while its dual has m-unknowns.

(3) The primal has n-unknowns while its dual has n-constraints.

(4) The n-coefficients of the objective function of primal (C_j) become the n-constant terms (b_i) of its dual.

(5) The m-constant terms of the primal (b_i) become the m-constant terms of the objective function (C_j) of its dual.

(6) The coefficients of the variables of the primal are transformed in their position in the dual. This means that the first column of the coefficients in the primal becomes the first row in the dual, the second column becomes the second row and so on.

(7) The *n*-variables (X_n) of the primal are replaced by the *m* new variables (Y_m) of its dual. This change affects the system of restrictions as well as the objective function.

(8) The sign of the inequalities in the set of restrictions of the primal ($\leq$) is reversed in the set of restrictions in its dual ($\geq$). In other words, if the inequalities in the primal are of the type $\leq$, then, they are of $\geq$ type in the dual.

(9) The sign of the inequalities restricting the variable ($\geq X_j$) to non-negative values in the primal is equal to the inequality sign of the new variable ($\geq Y_j$)of its dual.

(10) For writing the dual of the given maximization problem, we should first ensure that all the constraint inequalities are of the $\leq$ type and for writing the dual of the given minimization problem, the constraint inequalities should be of the $\geq$ type. We can see the application of these considerations with the help of given examples.

Example 4.10: Write the dual of the following primal LP problem.

Maximize,

$$Z = 2X_1 + 3X_2$$

Subject to,

$$2X_1 + X_2 \leq 20$$

and $$X_1 + 2X_2 \leq 20$$

$$X_1, X_2 \geq 0$$

Solution: The dual of the primal problem will be as under:

Minimize,

$$Z_y = 20Y_1 + 20Y_2$$

Subject to,

$$2Y_1 + Y_2 \geq 2$$

and $$Y_1 + 2Y_2 \geq 3$$

$$Y_1 \geq 0;\ Y_2 \geq 0$$

Example 4.11: Obtain the dual of the following LP problem.

Maximize,

$$Z_x = 20X_1 + 30X_2 + 10X_3$$

Subject to,

$$X_1 - X_3 \leq 4$$

$$12X_1 + 8X_2 \leq 15$$

$$X_1 + X_2 + X_3 \geq 8$$

$$4X_1 + 3X_2 - X_3 = 10$$
$$X_1 \geq 0; X_2 \geq 0; X_3 \geq 0$$

Solution: The dual of this problem will be as under:

NOTES

Mimimize,

$$Z_y = 4Y_1 + 15Y_2 - 8Y_3 + 10Y_4 - 10Y_5$$

Subject to,

$$Y_1 + 12Y_2 - Y_3 + 4Y_4 - 4Y_5 \geq 20$$
$$0Y_1 + 8Y_2 - Y_3 + 3Y_4 - 3Y_5 \geq 30$$
$$-1Y_1 + 0Y_2 - Y_3 - Y_4 + Y_5 \geq 10$$

and $\quad Y_1, Y_2, Y_3, Y_4, Y_5 \geq 0$

Since the above form of dual contains 5 variables in the objective function which are only four constraints in the given primal problem. This situation can be tackled if we take Y_6 as the difference between two non-negative variables Y_4 and be Y_3 (i.e., $Y = Y_4 - Y_5$). Accordingly, the above dual can be written as under:

Minimize,

$$Z_y = 4Y_1 + 15Y_2 - 8Y_3 + 10Y_6$$

Subject to,

$$Y_1 + 12Y_2 - Y_3 + 4Y_6 \geq 20$$
$$8Y_2 - Y_3 + 3Y_6 \geq 30$$
$$-Y_1 - Y_3 - Y_6 \geq 10$$

and

$Y_1, Y_2, Y_3 \geq 0$ and Y_6 unrestricted in sign.

Note: Whenever a constraint in the given problem has an equality sign its corresponding dual variable shall be unrestricted in sign. Similarly if there is unrestricted variable on the given primal problem its corresponding constraint shall have the = sign.

Example 4.12: Write the dual of the following LP problem.

Maximize,

$$Z_x = 5X_1 + 7X_2 + 9X_3$$

Subject to,

$$2X_1 + 2X_2 + 5X_3 \leq 12$$
$$6X_1 - 2X_2 + 4X_3 \geq 15$$

and $X_1, X_2 > 0$ and X_3 unrestricted in sign.

Solution: The dual of the given problem will be written as under:

Minimize,

$$Z_Y = 12Y_1 - 15Y_2$$

Subject to, $\quad 2Y_1 - 6Y_2 \geq 5$

$$2Y_1 + 2Y_2 \geq 7$$
$$5Y_1 - 4Y_2 = 9$$

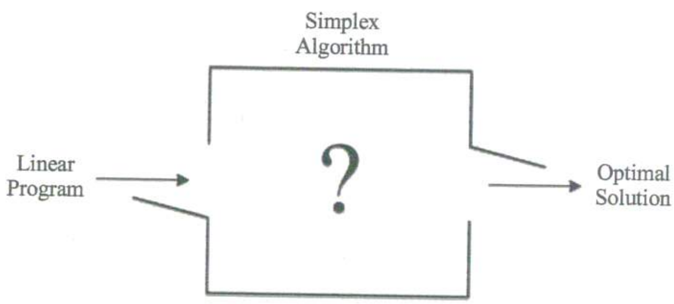
and $\quad Y_1 \geq 0; Y_2 \geq 0$

Since, in the given primal problem, X_3 is unrestricted in sign, we have taken $X_3 = (X_4 - X_5)$ and then using it the dual has been developed as above. It is instructive to note that the simplex method automatically identifies the dual basic solution. The optimal value of the objective function remains same as in the primal problem. Given an optimal solution of the primal problem, the *i*th dual variable acquires the coefficient of the *i*th slack variable in the optimal objective function equation as its optimal value. Thus, it is possible to identify the dual solution from the primal solution.

NOTES

4.7.1 Sensitivity AnalysisSensitivity Analysis

The optimal solution of a linear programming problem is formulated using various methods. You have learned the use and importance of dual variables to solve the LPP. Here, you will learn how sensitivity analysis helps to solve repeatedly the real problem in a little different form. Generally, these scenarios crop up as an end result of parameter changes due to the involvement of new advanced technologies and the accessibility of well-organized latest information for key (input) parameters or the 'what-if' questions. Thus, **sensitivity analysis** helps to produce optimal solution of simple perturbations for the key parameters. For optimal solutions, consider the simplex algorithm as a 'black box' which accepts the input key parameters to solve LPP as shown below:



Example 4.13: Illustrate sensitivity analysis using simplex method to solve the following LPP:

Maximize $Z = 20x_1 + 10x_2$

Subject to,

And $x_1, x_2 \geq 0$

Solution: Sensitivity analysis is done after making the initial and final tableau using the simplex method. Add slack variables to convert it into equation form.

Maximize $Z = 20x_1 + 10x_2 + 0S_1 + 0S_2$

Subject to:

Where $x_1, x_2 \geq 0$

To find basic feasible solution, we put $x_1 = 0$ and $x_2 = 0$. This gives $Z = 0$, $S_1 = 3$ and $S_2 = 7$. The initial table will be as follows:

NOTES

Initial table

		C_j	20	10	0	0	
C_B	B	X_B	X_1	X_2	S_1	S_2	Min $\frac{X_B}{X_i}$
0	S_1	3	1	1	1	0	3/1 = 3
←0	S_2	7	(3)	1	0	1	7/3 = 2.33
	Z_j	0	0	0	0	0	
	$Z_j - C_j$		−20↑	−10	0	0	

Find $\frac{X_B}{X_i}$ for each row and also find minimum for the second row. Here, $Z_j - C_j$ is maximum negative (–20). Hence, x_1 enters the basis and S_2 leaves the basis. It is shown with the help of arrows.

Key element is 3, key row is second row and key column is x_1. Now convert the key element into entering key by dividing each element of the key row by key element using the following formula:

$$\text{New element} = \text{Old element} - \left[\frac{\text{Product of elements in the key row and key column}}{\text{Key element}}\right]$$

The following is the first iteration table:

		C_j	20	10	0	0	
C_B	B	X_B	X_1	X_2	S_1	S_2	
←0	S_1	2/3	0	(2/3)	1	−1/3	$\frac{2}{3}/\frac{2}{3}=1$
20	X_1	7/3	1	1/3	0	1	$\frac{7}{3}/\frac{1}{3}=7$
	Z_j	140/3	20	20/3	0	20	
	$Z_j - C_j$	–	0	−10/3↑	0	20	

Since $Z_j - C_j$ has one value less than zero, that is, negative value hence, this is not yet the optima solution. Value –10/3 is negative; hence, x_2 enters the basis and S_1 leaves the basis. Key row is upper row.

		C_j	20	10	0	0
C_B	B	X_B	X_1	X_2	S_1	S_2
10	X_2	1	0	1	3/2	−1/2
20	X_1	4/3	1	0	0	4/3
	Z_j	110/3	20	10	0	25
	$Z_j - C_j$		0	0	0	25

$Z_j - C_j$ ³ 0 for all; hence, optimal solution is reached, where, $x_1 = \frac{4}{3}, x_2 = 1,$

$Z = \frac{80}{3} + 10 = \frac{110}{3}$.

Check Your Progress

18. What is duality?
19. What is primal?
20. How does a sensitivity analysis help in producing optimal solution?

4.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Linear programming is a decision-making technique under a set of given constraints and is based on the assumption that the relationships amongst the variables representing different phenomena are linear.
2. Criterion function is objective function which states the determinants of the quantity, to be either maximized or minimized.
3. Linear programming finds application in agricultural and various industrial problems.
4. Constraints are limitations under which planning is decided, these are restrictions imposed on decision variables.
5. Solution of a linear programming is a set of values $X_1, X_2, ..., X_n$, satisfying the constraints of the LPP.
6. In a given system of m linear equations with n variables ($m < n$), any solution which is obtained by solving m variables keeping the remaining $n - m$ variables zero is called a basic solution.
7. In a given system of m linear equations with n variables ($m < n$), where m variables are solved keeping remaining $n - m$ variables zero, m variables are called basic variables and the remaining variables are called non-basic variables.
8. Basic feasible solution is a basic solution which satisfies the condition in which all basic variables are non-negative.
9. An objective function is maximized when it is a profit function. It is minimized when it is a cost function.
10. Feasible solution of a LPP is a solution that satisfies the non-negativity restrictions of the LPP.
11. Feasible region is the region comprising all feasible solutions.
12. Optimal solution of a LPP is a feasible solution which optimizes (minimizes or maximizes) the objective function of the LPP.
13. Non-degenerate and degenerate solutions are the basic feasible solutions. In a problem which has exactly m positive variables, X_i ($i = 1, 2, ..., m$), i.e., none of the basic variables is zero, then it is called non-degenerate type and if one or more basic variables are zero then such basic feasible solution is said to be degenerate type.

NOTES

14. Simplex method is an iterative procedure for solving LPP in a finite number of steps. This method provides an algorithm which consists of moving from one vertex of the region of feasible solution to another in such a manner that the value of the objective function at the succeeding vertex is less or more as the case may be that at the previous vertex.
15. The leaving element is converted to unity by dividing the key equation by the key element and all other elements in its column to zero by using the formula:

    New element

    $$= \text{Old element} - \left[\frac{\text{Product of elements in key row and key column}}{\text{Key element}}\right]$$
16. By introducing slack variable, the problem is converted into standard form.
17. M method is used to find the starting basic feasible solution each time it exists when an infeasible starting basic solution is given.
18. For every given linear programming problem there is another intimately related linear programming problem referred to as its dual. The duality theorem states that 'for every maximization (or minimization) problem in linear programming there is a unique similar problem of minimization (or maximization) involving the same date which describes the original problem'.
19. The original problem is referred to as the 'primal'. The 'dual' of a dual problem is the primal. Thus the primal and dual problems are replicas of each other.
20. The sensitivity analysis helps to produce optimal solution of simple perturbations for the key parameters. For optimal solutions, consider the simplex algorithm as a 'black box' which accepts the input key parameters to solve LPP.

4.9 SUMMARY

- Linear Programming (LP) is a major innovation since World War II in the field of business decision-making, particularly under conditions of certainty.
- The word 'Linear' means that the relationships are represented by straight lines, i.e., the relationships are of the form $y = a + bx$ and the word 'Programming' means taking decisions systematically.
- LP is generally used in solving maximization (sales or profit maximization) or minimization (cost minimization) problems subject to certain assumptions.
- LP is extensively used in solving resource allocation problems.
- The term linearity implies straight line or proportional relationships among the relevant variables.
- Process means the combination of particular inputs to produce a particular output.

- Criterion function is also known as objective function which states the determinants of the quantity either to be maximized or to be minimized.
- Feasible solutions are all those possible solutions which can be worked upon under given constraints. The region comprising of all feasible solutions is referred as feasible region.
- Optimum solution is the best of the feasible solutions. Z denotes the optimal value of the objective function for a given system.
- LP model is based on the assumptions of proportionality, additivity, certainty, continuity and finite choices.
- The applications of linear programming problems are based on linear programming matrix coefficients and data transmission prior to solving the simplex algorithm.
- The portfolio optimization template calculates the optimal capital of investments that gives the highest return for the least risk.
- The objective function, the set of constraints and the non-negative constraints together form a linear programming problem.
- Simple linear programming problem with two decision variables can be easily solved by graphical method.
- A set of values $X_1, X_2, ..., X_n$ which satisfies the constraints of the LPP is called its solution.
- Given a system of m linear equations with n variables ($m < n$), any solution which is obtained by solving m variables keeping the remaining $n - m$ variables zero is called a basic solution. Such m variables are called basic variables and the remaining variables are called non-basic variables.
- A basic feasible solution is a basic solution which also satisfies all basic variables which are non-negative.
- If the value of the objective function Z can be increased or decreased indefinitely, such solutions are called unbounded solutions.
- Simplex method is an iterative procedure for solving LPP in a finite number of steps.
- For the solution of any LPP by simplex algorithm, the existence of an initial basic feasible solution is always assumed.
- In simplex algorithm, the M Method is used to deal with the situation where an infeasible starting basic solution is given.
- For every given linear programming problem there is another intimately related linear programming problem referred to as its dual.
- The duality theorem states that 'for every maximization (or minimization) problem in linear programming there is a unique similar problem of minimization (or maximization) involving the same date which describes the original problem'. The original problem is referred to as the 'primal'. The 'dual' of a dual problem is the primal. Thus the primal and dual problems are replicas of each other.

NOTES

4.10 KEY TERMS

- **Linear programming:** It is a decision-making technique under a set of given constraints and is based on the assumption that the relationships amongst the variables representing different phenomena are linear.
- **Decision variables:** These are the variables that form objective function and on which the cost or profit depends.
- **Linearity:** It is the line straight or proportional relationships among the relevant variables. Linearity in economic theory is known as constant return.
- **Process:** It is the combination of one or more inputs to produce a particular output.
- **Criterion function:** It is an objective function which states the determinants of the quantity to be either maximized or minimized.
- **Constraints:** It refers to limitations under which planning is decided. Restrictions imposed on decision variables.
- **Feasible solution:** It refers to any solution to a LPP which satisfies the non-negativity restrictions of the LPP.
- **Feasible region:** It is the region comprising all feasible solutions.
- **Optimal solution:** It refers to any feasible solution which optimizes (minimizes or maximizes) the objective function of the LPP.
- **Proportionality:** It is an assumption made in the objective function and constraint inequalities. In economic terminology this means that there are constant returns to scale.
- **Certainty:** It refers to the assumption that includes prior knowledge of all the coefficients in the objective function, the coefficients of the constraints and the resource values. LP model operates only under conditions of certainty.
- **Additivity:** It is an assumption which means that the total of all the activities is given by the sum total of each activity conducted separately.
- **Continuity:** It is an assumption which means that the decision variables are continuous.
- **Finite choices:** It refers to an assumption that implies that finite numbers of choices are available to a decision-maker and the decision variables do not assume negative values.
- **Solution:** It is s set of values $X_1, X_2, ..., X_n$ which satisfies the constraints of the LPP.
- **Basic solution:** In a given system of m linear equations with n variables ($m < n$), any solution which is obtained by solving m variables keeping the remaining $n - m$ variables zero is called a basic solution.
- **Basic feasible solution:** It is a basic solution which also satisfies the condition in which all basic variables are non-negative.

- **Canonical form:** It is irrespective of the objective function. All the constraints are expressed as equations and right hand side of each constraint and all variables are non-negative.
- **Slack variables:** If the constraints of a general LPP be given as $\Sigma a_{ij} X_i \leq b_i$ $(i = 1, 2, ..., m; j = 1, 2, ..., n)$, then the non-negative variables S_i is introduced to convert the inequalities '≤' to the equalities are called slack variables.
- **Surplus variables:** If the constraints of a general LPP be $\Sigma a_{ij} X_i \geq b_i$ $(i = 1, 2, ..., m; j = 1, 2, ..., n)$, then non-negative variables S_i introduced to convert the inequalities '≥' to the equalities are called surplus variables.

4.11 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is meant by proportionality in linear programming?
2. What do you understand by certainty in linear programming?
3. What is meant by continuity in linear programming?
4. What are finite choices in the context of linear programming?
5. What are the basic constituents of an LP model?
6. What is the canonical form of a LPP?
7. What are characteristics of the canonical form?
8. What are slack variables? Where are they used?
9. What do you understand by surplus variables?
10. What is the simplex method?
11. What is the importance of the *M* method?
12. What is dual?
13. How will you transform a given primal problem into a dual problem?

Long-Answer Questions

1. A company manufactures 3 products *A*, *B* and C. The profits are: ₹ 3, ₹ 2 and ₹ 4 respectively. The company has two machines and given below is the required processing time in minutes for each machine on each product.

	Products		
Machines	*A*	*B*	*C*
I	4	3	5
II	2	2	4

Machines I and II have 2000 and 2500 minutes respectively. The company must manufacturers 100 *A*'s 200 *B*'s and 50 *C*'s but no more than 150 *A*'s. Find the number of units of each product to be manufactured by the company to maximize the profit. Formulate the above as a LP Model.

NOTES

2. A company produces two types of leather belts A and B. A is of superior quality and B is of inferior quality. The respective profits are ₹ 10 and ₹ 5 per belt. The supply of raw material is sufficient for making 850 belts per day. For belt A, a special type of buckle is required and 500 are available per day. There are 700 buckles available for belt B per day. Belt A needs twice as much time as that required for belt B and the company can produce 500 belts if all of them were of the type A. Formulate a LP Model for the given problem.
3. The standard weight of a special purpose brick is 5 kg and it contains two ingredients B_1 and B_2, where B_1 costs ₹ 5 per kg and B_2 costs ₹ 8 per kg. Strength considerations dictate that the brick contains not more than 4 kg of B_1 and a minimum of 2 kg of B_2 since the demand for the product is likely to be related to the price of the brick. Formulate the given problem as a LP Model.
4. Egg contains 6 units of vitamin A per gram and 7 units of vitamin B per gram and 12 units of vitamin B per gram and costs 20 paise per gram. The daily minimum requirement of vitamin A and vitamin B are 100 units and 120 units respectively. Find the optimal product mix.
5. A company desires to devote the excess capacity of the three machines lathe, shaping machine and milling machine to make three products A, B and C. The available time per month in these machinery are tabulated below:

Machine	*Lathe*	*Shaping*	*Milling*
Available Time/Month	200 hrs	100 hrs	180 hrs

The time taken to produce each unit of the products A, B and C on the machines is displayed in the table below.

	Lathe	*Shaping*	*Milling*
Product A hrs	6	2	4
Product B hrs	2	2	–
Product C hrs	3	–	3

The profit per product would be ₹ 20, ₹ 16 and ₹ 12 respectively on product A, B and C.

Formulate a LPP to find the optimum product-mix.

6. Solve the following by graphical method:

(*i*) Max $Z = X_1 - 3X_2$

Subject to,
$$X_1 + X_2 \leq 300$$
$$X_1 - 2X_2 \leq 200$$
$$2X_1 + X_2 \leq 100$$
$$X_2 \leq 200$$
$$X_1, X_2 \geq 0$$

(*ii*) Max $Z = 5X + 8Y$

Subject to,
$$3X + 2Y \leq 36$$
$$X + 2Y \leq 20$$

$$3X + 4Y \le 42$$
$$X, Y \ge 0$$

(*iii*) Max $Z = X - 3Y$

Subject to, $X + Y \le 300$

$X - 2Y \le 200$

$X + Y \le 100$

$Y \ge 200$

and $X, Y \ge 0$

NOTES

7. Solve graphically the following LPP:

Max $Z = 20X_1 + 10X_2$

Subject to, $X_1 + 2X_2 \le 40$

$3X_1 + X_2 \ge 30$

$4X_1 + 3X_2 \ge 60$

and $X_1, X_2 \ge 0$

8. Using simplex method, find non-negative values of X_1, X_2 and X_3 when

(*i*) Max $Z = X_1 + 4X_2 + 5X_3$

Subject to the constraints,

$3X_1 + 6X_2 + 3X_3 \le 22$

$X_1 + 2X_2 + 3X_3 \le 14$ and

$3X_1 + 2X_2 \le 14$

(*ii*) Max $Z = X_1 + X_2 + 3X_3$

Subject to, $3X_1 + 2X_2 + X_3 \le 2$

$2X_1 + X_2 + 2X_3 \le 2$

$X_1, X_2, X_3 \ge 0$

(*iii*) Max $Z = 10X_1 + 6X_2$

Subject to, $X_1 + X_2 \le 2$

$2X_1 + X_2 \le 4$

$3X_1 + 8X_2 \le 12$

$X_1, X_2 \ge 0$

(*iv*) Max $Z = 30X_1 + 23X_2 + 29X_3$

Subject to the constraints,

$6X_1 + 5X_2 + 3X_3 \le 52$

$6X_1 + 2X_2 + 5X_3 \le 14$

$X_1, X_2, X_3 \ge 0$

(*v*) Max $Z = X_1 + 2X_2 + X_3$

Subject to, $2X_1 + X_2 - X_3 \ge -2$

$-2X_1 + X_2 - 5X_3 \le 6$

$4X_1 + X_2 + X_3 \le 6$

$X_1, X_2, X_3 \ge 0$

NOTES

9. Solve the following LPP applying M method:

Maximize, $Z = 3X_1 + 4X_2$

Subject to, $2X_1 + X_2 \leq 600$

$X_1 + X_2 \leq 225$

$5X_1 + 4X_2 \leq 1000$

$X_1 + 2X_2 \geq 150$

$X_1, X_2 \geq 0$

10. Write the dual of the following LP problem:

Maximize, $Z = 5X_1 + 6X_3$

Subject to, $X_1 + 2X_2 = 5$

$-X_1 + 5X_2 \geq 3$

$4X_1 + 7X_2 \leq 8$

and X_1 unrestricted in sign

$X_2 \geq 0$

11. Write the primal of the following dual:

Minimize, $ZY = 10Y_1 + 8Y_2$

Subject to, $Y_1 + 2Y_2 \geq 5$

$2Y_1 - Y_2 \geq 12$

$Y_1 + 3Y_2 \geq 4$

and $Y_1 \geq 0; \geq Y_2$ unrestricted in sign.

4.12 FURTHER READING

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House.

Gupta, S. C. 2006. *Fundamentals of Statistics*. New Delhi: Himalaya Publishing House.

Gupta, S. P., 2005. *Statistical Methods*. New Delhi: Sultan Chand and Sons.

Hooda, R. P. 2002. *Statistics for Business and Economics*. New Delhi: Macmillan India.

Kothari, C. R., 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Monga, G. S. 2000. *Mathematics and Statistics for Economics*. New Delhi: Vikas Publishing House

Gupta, S.P. 2006. *Statistical Methods*. New Delhi: S. Chand & Co. Ltd.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods*, 23rd edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Gupta, S.C. and V.K. Kapoor. *Fundamentals of Mathematical Statistics*. New Delhi: Sultan Chand & Sons.

Sharma, A. 2009. *Operations Research.* New Delhi: Himalaya Publishing House.

Singh, J. K. 2009. *Business Mathematics*. New Delhi: Himalaya Publishing House.

Kalavathy, S. 2010. *Operations Research with C Programmes,* 3rd edition. New Delhi: Vikas Publishing House Pvt. Ltd.

Kothari, C. R. 1992. *An Introduction to Operational Research.* New Delhi: Vikas Publishing House Pvt. Ltd.

NOTES

UNIT 5 TRANSPORTATION PROBLEM, ASSIGNMENT PROBLEMS, INTEGER AND GOAL PROGRAMMING

Structure

5.0 INTRODUCTION

The transportation problem is a subclass of a linear programming problem (LPP). Transportation problems deal with the objective of transporting various quantities of a single homogeneous commodity initially stored at various origins, to different destinations, in a way that keeps transportation cost at a minimum. The solution of any transportation problem is obtained in two stages, namely initial solution and optimal solution. There are three methods of obtaining an initial solution. These are: North West Corner Rule, Least Cost Method and Vogel's Approximation Method (VAM). VAM is preferred since the solution obtained this way is very close to the optimal solution. The optimal solution of any transportation problem is a feasible solution that minimizes the total cost. An optimal solution is the second stage of a solution obtained by improving the initial solution. Modified Distribution (MODI) method is used to obtain optimal solutions and optimality tests. This unit also deals with a very interesting method called the 'assignment technique', which is applicable to a class of very practical problems generally called 'assignment problems'. The objective of assignment problems is to assign a number of origins (jobs) to the equal number of destinations (persons) at a minimum cost or maximum profit. A linear programming problem in which all or some of the decision variables are constrained to assume non-negative integer values is called an Integer Programming Problem (IPP). In a linear programming problem, if all variables are required to take integral values then it is called the Pure (all) Integer Programming Problem (Pure IPP). If only some of the variables in the optimal solution of a LPP are restricted to assume non-negative integer values, while the remaining variables are free to take any non-negative values, then it is called a Mixed Integer Programming Problem (Mixed IPP). Further, if all the variables in the optimal solution are allowed to take values 0 or 1, then the problem is called the 0–1 Programming Problem or Standard Discrete Programming Problem. In this unit, you will learn about the applications of the transportation problem and the solutions and rules to solve such problems.

5.1 OBJECTIVES

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

- Discuss transportation models and their special cases
- State optimum solution using MODI method
- Describe assignment problem and its mathematical formulation
- Differentiate between transportation and assignment problems
- Describe unbalanced assignment problem and modified matrix
- Assess maximization in assignment problem
- Analyse the importance of integer programming problems
- Discuss the applications and types of integer programming problem
- Explain the concept of mixed integer programming problem
- Describe the Branch and Bound Method and the concept of goal programming

5.2 TRANSPORTATION PROBLEM

The transportation problem (TP) is one of the subclasses of LPP (Linear Programming Problem) in which the objective is to transport various quantities of a single homogeneous commodity that are initially stored at various origins to different destinations in such a way that the transportation cost is minimum. To achieve this objective, we must know the amount and location of available supplies and the quantities demanded. In addition, we must know the costs that result from transporting one unit of commodity from various origins to various destinations.

5.2.1 Formulation of Transportation Problem (TP)

Consider a transportation problem with *m* origins (rows) and *n* destinations (columns). Let C_{ij} be the cost of transporting one unit of the product from the *i*th origin to *j*th destination, a_i the quantity of commodity available at origin *i*, and b_j

the quantity of commodity needed at destination j. X_{ij} is the quantity transported from ith origin to jth destination. This transportation problem can be stated in the following tabular form.

NOTES

***Table 5.1** Transportation Problem (TP)*

Destinations

Origins	1	2	3	...	n	Capacity
1	C_{11} X_{11}	C_{12} X_{12}	C_{13} X_{13}	...	C_{1n} X_{1n}	a_1
2	C_{21} X_{21}	C_{22} X_{22}	C_{23} X_{23}	...	C_{2n} X_{2n}	a_2
3	C_{31} X_{31}	C_{32} X_{32}	C_{33}	...	C_{3n} X_{3n}	a_3
m	C_{m1} X_{m1}	C_{m2} X_{m2}	C_{m3} X_{m3}	...	C_{mn} X_{mn}	a_m
Demand	b_1	b_2	b_3	...	b_n	
						$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

The linear programming model representing the transportation problem is given by,

$$\text{Minimize } Z = \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}X_{ij}$$

Subject to the constraints,

$$\sum_{j=1}^{n} X_{ij} = a_i \qquad i = 1, 2, ..., n$$

(Row Sum)

$$\sum_{i=1}^{m} X_{ij} = b_j \qquad j = 1, 2, ..., n$$

(Column Sum)

$$X_{ij} \geq 0 \quad \text{For all } i \text{ and } j$$

The given transportation problem is said to be balanced if,

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$$

that is, the total supply is equal to the total demand.

5.2.2 Initial Basic Feasible Solution

Feasible solution: Any set of non-negative allocations (X_{ij}>0) which satisfies the row and column sum (rim requirement) is called a feasible solution.

Basic feasible solution: A feasible solution is called a basic feasible solution if the number of non-negative allocations is equal to $m + n - 1$, where m is the number of rows and n the number of columns in a transportation table.

Non-degenerate basic feasible solution: Any feasible solution to a transportation problem containing m origins and n destinations is said to be non-degenerate, if it contains $m + n - 1$ occupied cells and each allocation is in independent positions.

The allocations are said to be in independent positions if it is impossible to form a closed path. Closed path means by allowing horizontal and vertical lines and when all the corner cells are occupied.

The allocations in the following tables are not in independent positions.

	*	*
	*	*

*		*
*		*

	*	*	
	*		
	*	*	

The allocations in the following tables are in independent positions.

	*	
*	*	*
*		

*	*		
	*		*
		*	*

Degenerate basic feasible solution: If a basic feasible solution contains less than $m + n - 1$ non-negative allocations, it is said to be degenerate.

5.2.3 Moving Towards Optimality

An **optimal solution** is a feasible solution (not necessarily basic) which minimizes the total cost.

The solution of a transportation problem (TP) can be obtained in two stages, namely initial solution and optimum solution.

Initial solution can be obtained by using any one of the three methods, namely:

- North West Corner Rule (NWCR)
- Least Cost Method or Matrix Minima Method
- Vogel's Approximation Method (VAM)

VAM is preferred over the other two methods, since the initial basic feasible solution obtained by this method is either optimal or very close to the optimal solution.

The cells in the transportation table can be classified into occupied cells and unoccupied cells. The allocated cells in the transportation table are called *occupied cells* and the *empty cells* in the transportation table are called *unoccupied cells*.

The improved solution of the initial basic feasible solution is called optimal solution which is the second stage of solution, that can be obtained by MODI.

North West Corner Rule

The following steps explain the North West Corner Rule:

Step 1: Starting with the cell at the upper left corner (North West) of the transportation matrix, we allocate as much as possible so that either the capacity of the first row is exhausted or the destination requirement of the first column is satisfied, that is, $X_{11} = \text{Min}\ (a_1, b_1)$.

Step 2: If $b_1 > a_1$, we move down vertically to the second row and make the second allocation of magnitude $X_{22} = \text{Min}(a_2\ b_1 - X_{11})$ in the cell (2, 1).

If $b_1 < a_1$, move right horizontally to the second column and make the second allocation of magnitude $X_{12} = \text{Min}\ (a_1, X_{11} - b_1)$ in the cell (1, 2).

If $b_1 = a_1$, there is a tie for the second allocation. We make the second allocations of magnitude,

$$X_{12} = \text{Min}\ (a_1 - a_1, b_1) = 0 \text{ in the cell } (1,2)$$

$$\text{or, } X_{21} = \text{Min}\ (a_2, b_1 - b_1) = 0 \text{ in the cell } (2,1)$$

Step 3: Repeat steps 1 and 2 moving down towards the lower right corner of the transportation table until all the rim requirements are satisfied.

Example 5.1: Obtain the initial basic feasible solution of a transportation problem whose cost and rim requirement table is as follows:

Origin\Destination	D_1	D_2	D_3	*Supply*
O_1	2	7	4	5
O_2	3	3	1	8
O_3	5	4	7	7
O_4	1	6	2	14
Demand	7	9	18	34

Solution: Since $\Sigma a_i = 34 = \Sigma b_j$, there exists a feasible solution to the transportation problem. We obtain the initial feasible solution as follows:

The first allocation is made in the cell (1, 1), the magnitude being $X_{11} = \text{Min}\ (5, 7) = 5$.

The second allocation is made in the cell (2, 1) and the magnitude of the allocation is given by $X_{21} = \text{Min}\ (8, 7 - 5) = 2$.

	D_1	D_2	D_3	*Supply*
O_1	2 ⑤	7	4	~~5~~ 0
O_2	3 ②	3 ⑥	1	~~8~~ ~~6~~ 0
O_3	5	4 ③	7 ④	~~7~~ ~~4~~ 0
O_4	1	6	2 (14)	~~14~~ 0
Demand	~~7~~ ~~2~~ 0	~~9~~ ~~3~~ 0	~~18~~ ~~14~~ 0	34

The third allocation is made in the cell (2, 2) the magnitude $X_{22} = \text{Min}\ (8 - 2, 9) = 6$.

The magnitude of the fourth allocation is made in the cell (3, 2) given by X_{32} = Min (7, 9 – 6) = 3.

The fifth allocation is made in the cell (3, 3) with magnitude X_{33} = Min (7 – 3, 14) = 4.

The final allocation is made in the cell (4, 3) with magnitude X_{43} = Min (14, 18 – 4) = 14.

Hence, we get the initial basic feasible solution to the given TP and is given by,

$X_{11} = 5; X_{21} = 2; X_{22} = 6; X_{32} = 3; X_{33} = 4; X_{43} = 14$

Total Cost = 2 × 5 + 3 × 2 + 3 × 6 + 3 × 4 + 4 × 7 + 2 × 14

=10 + 6 + 18 + 12 + 28 + 28 = ₹ 102

Example 5.2: Determine an initial basic feasible solution to the following transportation problem using North West Corner Rule.

	D_1	D_2	D_3	D_4	*Supply*
O_1	6	4	1	5	14
O_2	8	9	2	7	16
O_3	4	3	6	2	5
Required	6	10	15	4	35

Solution: The problem is a balanced TP, as the total supply is equal to the total demand. Using the steps, we find the initial basic feasible solution as given in the following table.

	D_1	D_2	D_3	D_4	*Supply*
O_1	[6] (6)	[4] (8)	[1]	[5]	~~14~~ ~~8~~ 0
O_2	[8]	[9] (2)	[2] (14)	[7]	~~16~~ ~~14~~ 0
O_3	[4]	[3]	[6] (1)	[2] (4)	~~5~~ ~~4~~
Demand	~~6~~	~~10~~ ~~2~~ 0	~~15~~ ~~1~~ 0	~~4~~	35

The solution is given by,

$X_{11} = 6; X_{12} = 8; X_{22} = 2; X_{23} = 14; X_{33} = 1; X_{34} = 4$

Total Cost = 6 × 6 + 4 × 8 + 2 × 9 + 2 × 14 +6 × 1 + 2 × 4

= 36 + 32 + 18 + 28 + 6 + 8 = ₹ 128

Least Cost or Matrix Minima Method

The following methods discuss the Least Cost or Matrix Minima Method:

Step 1 : Determine the smallest cost in the cost matrix of the transportation table. Let it be C_{ij}. Allocate X_{ij} = Min (a_i, b_j) in the cell (i, j).

Step 2: If $X_{ij} = a_i$ cross off the ith row of the transportation table and decrease b_j by a_i. Then go to Step 3.

If $X_{ij} = b_j$ cross off the jth column of the transportation table and decrease a_i by b_j. Go to Step 3.

If $X_{ij} = a_i = b_j$ cross off either the ith row or the jth column, but not both.

Step 3: Repeat steps 1 and 2 for the resulting reduced transportation table until all the rim requirements are satisfied. Whenever the minimum cost is not unique, make an arbitrary choice among the minima.

Example 5.3: Obtain an initial feasible solution to the following TP using Matrix Minima Method.

	D_1	D_2	D_3	D_4	*Supply*
O_1	1	2	3	4	6
O_2	4	3	2	0	8
O_3	0	2	2	1	10
Demand	4	6	8	6	24

Solution: Since $\Sigma a_i = \Sigma b_j = 24$, there exists a feasible solution to the TP using the steps in the least cost method, the first allocation is made in the cell (3, 1) the magnitude being $X_{31} = 4$. This satisfies the demand at the destination D_1 and we delete this column from the table as it is exhausted.

	D_1	D_2	D_3	D_4	*Supply*
O_1	1	2 (6)	3	4	~~6~~ 0
O_2	4	3	2 (2)	0 (6)	~~8~~ ~~2~~
O_3	0 (4)	2	2 (6)	1	~~10~~ ~~6~~
Demand	~~4~~ 0	~~6~~ 0	~~8~~ ~~2~~ 0	~~6~~ 0	24

The second allocation is made in the cell (2, 4) with magnitude X_{24} = Min (6, 8) = 6. Since it satisfies the demand at the destination D_4, it is deleted from the table. From the reduced table, the third allocation is made in the cell (3, 3) with magnitude X_{33} = Min (8, 6) = 6. The next allocation is made in the cell (2, 3) with magnitude X_{23} of Min (2, 2) = 2. Finally, the allocation is made in the cell (1, 2) with magnitude X_{12} = Min (6, 6) = 6. Now, all the requirements have been satisfied and hence, the initial feasible solution is obtained.

The solution is given by,

$$X_{12} = 6; X_{23} = 2; X_{24} = 6; X_{31} = 4; X_{33} = 6$$

Since the total number of occupied cells = 5 < $m + n - 1$

We get a degenerate solution.

Total cost = $6 \times 2 + 2 \times 2 + 6 \times 0 + 4 \times 0 + 6 \times 2$

= 12 + 4 + 12 = ₹ 28

Example 5.4: Determine an initial basic feasible solution for the following TP, using the Least Cost Method.

	D_1	D_2	D_3	D_4	*Supply*
O_1	6	4	1	5	14
O_2	8	9	2	7	16
O_3	4	3	6	2	5
Demand	6	10	15	4	35

Solution: Since $\Sigma a_i = \Sigma b_j$, there exists a basic feasible solution. Using the steps in least cost method, we make the first allocation to the cell (1, 3) with magnitude X_{13} = Min (14, 15) = 14 (as it is the cell having the least cost).

This allocation exhausts the first row supply. Hence, the first row is deleted. From the reduced table, the next allocation is made in the next least cost cell (2, 3) which is chosen arbitrarily with magnitude X_{23} = Min (1, 16) = 1. This exhausts the third column destination.

From the reduced table, the next least cost cell is (3, 4) for which allocation is made with magnitude Min (4, 5) = 4. This exhausts the destination D_4 requirement. Delete this fourth column from the table. The next allocation is made in the cell (3, 2) with magnitude X_{32} = Min (1, 10) = 1 which exhausts the third origin capacity. Hence, the third row is exhausted. From the reduced table, the next allocation is given to the cell (2,1) with magnitude X_{21} = Min (6, 15) = 6. This exhausts the first column requirement. Hence, it is deleted from the table.

Finally, the allocation is made to the cell (2, 2) with magnitude X_{22} = Min (9, 9) = 9 which satisfies the rim requirement. These allocations are shown in the following transportation table:

	D_1	D_2	D_3	D_4	*Supply*
O_1	6	4	1 (14)	5	~~14~~
O_2	8	9	2	7	16
O_3	4	3	6	2	5
Demand	6	10	~~15~~ 1	4	

(I Allocation)

	D_1	D_2	D_3	D_4	*Supply*
O_2	8	9	2 (1)	7	~~16~~ 15
O_3	4	3	6	2	5
Demand	6	10	~~1~~ 0	4	

(II Allocation)

	D_1	D_2	D_4	*Supply*
O_2	8	9	7	15
O_3	4	3	2 (4)	~~5~~ 1
Demand	6	10	~~4~~ 0	

(III Allocation)

	D_1	D_2	*Supply*
O_2	8	9	15
O_3	4	3 (1)	~~1~~ 0
Demand	6	~~10~~ 9	

(IV Allocation)

	D_1	D_2	*Supply*
O_2	8 (6)	9 (9)	15
Demand	6	9	

(V, VI Allocation)

The following table gives the initial basic feasible solution.

	D_1	D_2	D_3	D_4	*Supply*
O_1	6	4	1 (14)	5	14
O_2	8 (6)	9 (9)	2 (1)	7	16
O_3	4	3 (1)	6	2 (4)	5
Demand	6	10	15	4	

The solution is given by,

$X_{13}=14; X_{21}=6; X_{22}=9; X_{23}=1; X_{32}=1; X_{34}=4$

Transportation cost,

$= 14 \times 1 + 6 \times 8 + 9 \times 9 + 1 \times 2 + 3 \times 1 + 4 \times 2$

$= 14 + 48 + 81 + 2 + 3 + 8 = ₹\ 156$

Vogel's Approximation Method (VAM)

The steps involved in Vogel's Approximation Method (VAM) for finding the initial solution are as follows:

Step 1: Find the penalty cost, namely the difference between the smallest and the next smallest costs in each row and column.

Step 2: Among the penalties as found in Step (1), choose the maximum penalty. If this maximum penalty is more than one (that is, if there is a tie), choose any one arbitrarily.

Step 3: In the selected row or column as by Step (2), find out the cell having the least cost. Allocate to this cell as much as possible depending on the capacity and requirements.

Step 4: Delete the row or column which is fully exhausted. Again, compute the column and row penalties for the reduced transportation table and then go to Step (2). Repeat the procedure until all the rim requirements are satisfied.

Note: If the column is exhausted, then there is a change in row penalty and vice versa.

Example 5.5: Find the initial basic feasible solution for the following transportation problem using VAM.

		Destination				
		D_1	D_2	D_3	D_4	*Supply*
Origin	O_1	11	13	17	14	250
	O_2	16	18	14	10	300
	O_3	21	24	13	10	400
	Demand	200	225	275	250	950

Solution: Since $\Sigma a_i = \Sigma b_j = 950$, the problem is balanced and there exists a feasible solution to the problem.

First, we find the row and column penalty P_I as the difference between the least and the next least cost. The maximum penalty is 5. Choose the first column arbitrarily. In this column, choose the cell having the least cost name (1, 1). Allocate to this cell with minimum magnitude (that is, Min (250, 200) = 200). This exhausts the first column. Delete this column. Since a column is deleted, there is a change in row penalty P_{II} while column penalty P_{II} remains the same. Continuing in this manner, we get the remaining allocations as given in the following table:

NOTES

I Allocation

	D_1	D_2	D_3	D_4	Supply	P_I
O_1	11 (200)	13	17	14	50 ~~250~~	2
O_2	16	18	14	10	300	4
O_3	21	24	13	10	400	3
Demand	~~200~~ 0	225	275	250		
P_I	5↑	5	(1)	0		

II Allocation

	D_2	D_3	D_4	Supply	P_{II}
O_1	13 (50)	17	14	~~50~~	(1)
O_2	18	14	10	300	4
O_3	24	13	10	400	3
Demand	~~225~~ 175	275	250		
P_{II}	5↑	(1)	0		

III Allocation

	D_2	D_3	D_4	Supply	P_{III}
O_2	18 (175)	14	10	~~300~~ 125	4
O_3	24	13	10	400	3
Demand	~~175~~ 0	275	250		
P_{III}	6↑	1	0		

IV Allocation

	D_3	D_4	Supply	P_{IV}
O_2	14	10 (125)	~~125~~ 0	4 ←
O_3	13	10	400	3
Demand	275	~~250~~ 125		
P_{IV}	1	0		

V Allocation

	D_3	D_4	Supply	P_V
O_3	13 (275)	10	~~400~~ 125	3
Demand	~~275~~ 0	125		
P_V	13↑	10		

VI Allocation

	D_4	Supply	P_{VI}
O_3	10 (125)	~~125~~ 0	10 ←
Demand	~~125~~ 0		
P_{VI}	10		

Finally, we arrive at the initial basic feasible solution which is shown in the following table:

	D_1	D_2	D_3	D_4	Supply
O_1	11 (200)	13 (50)	17	14	250
O_2	16	18 (175)	14	10 (125)	300
O_3	21	24	13 (275)	10 (125)	400
Demand	200	225	275	250	

There are six positive independent allocations which equals to $m+n-1 = 3+4-1$. This ensures that the solution is a non-degenerate basic feasible solution.

Transportation cost

$= 11 \times 200 + 13 \times 50 + 18 \times 175 + 10 \times 125 + 13 \times 275 + 10 \times 125$

$=$ ₹ 12,075

Example 5.6: Find the initial solution to the following TP using VAM.

		Destination				
		D_1	D_2	D_3	D_4	*Supply*
Factory	F_1	3	3	4	1	100
	F_2	4	2	4	2	125
	F_3	1	5	3	2	75
	Demand	120	80	75	25	300

Solution: Since $\Sigma a_i = \Sigma b_j$, the problem is a balance TP. Hence, there exists a feasible solution.

	D_1	D_2	D_3	D_4	*Supply*	P_I	P_{II}	P_{III}	P_{IV}	P_V	P_{VI}
F_1	3 (45)	3	4 (30)	1 (25)	100	2	2 ←	0	1 ←	4	4
F_2	4	2 (80)	4 (45)	2	125	0	0	2 ←	0	4 ←	
F_3	1 (75)	5	3	2	75	1		←			
Demand	120	80	75	25							
P_I	2↑	1	1	1							
P_{II}	1	1	0	1							
P_{III}	1	1	0								
P_{IV}	1		0								
P_V			0								
P_{VI}			4↑								

Finally, we have the initial basic feasible solution as given in the following table.

	D_1	D_2	D_3	D_4	*Supply*
F_1	3 (45)	3	4 (30)	1 (25)	100
F_2	4	2 (80)	4 (45)	2	125
F_3	1 (75)	5	3	2	75
Demand	120	80	75	25	

There are six independent non-negative allocations equal to $m+n-1=3+4-1=6$. This ensures that the solution is non-degenerate basic feasible.

Transportation cost

$= 3 \times 45 + 4 \times 30 + 1\times 25 + 2 \times 80 + 4 \times 45 + 1 + 75$

$= 135 + 120 + 25 + 160 + 180 + 75$

$=$ ₹ 695

5.2.4 Transportation Algorithm (MODI) Method

In this section, we will discuss the Optimum Solution Using MODI Method.

Optimality Test

Once the initial basic feasible solution has been computed, the next step in the problem is to determine whether the solution obtained is optimum or not.

Optimality test can be conducted to any initial basic feasible solution of a TP, provided such allocations has exactly $m + n - 1$ non-negative allocations, where m is the number of origins and n is the number of destinations. Also, these allocations must be in independent positions.

To perform this optimality test, we shall discuss the modified distribution method (MODI). The various steps involved in the MODI method for performing optimality test are as follows.

MODI Method

Step 1: Find the initial basic feasible solution of a TP by using any one of the three methods.

Step 2: Find out a set of numbers u_i and v_j for each row and column satisfying $u_i + v_j = C_{ij}$ for each occupied cell. To start with, we assign a number '0' to any row or column having the maximum number of allocations. If this maximum number of allocations is more than 1, choose any one arbitrarily.

Step 3: For each empty (unoccupied) cell, we find the sum u_i and v_j written in the bottom left corner of that cell.

Step 4: Find out for each empty cell the net evaluation value $\Delta_{ij} = C_{ij} - (u_i + v_j)$ and which is written at the bottom right corner of that cell. This step gives the optimality conclusion as follows:

- If all $\Delta_{ij} > 0$ (i.e., all the net evaluation values), the solution is optimum and a unique solution exists.
- If $\Delta_{ij} \geq 0$, then the solution is optimum, but an alternate solution exists.
- If at least one $\Delta_{ij} < 0$, the solution is not optimum. In this case, we go to the next step to improve the total transportation cost.

Step 5: Select the empty cell having the most negative value of Δ_{ij}. From this cell, we draw a closed path by drawing horizontal and vertical lines with the corner cells occupied. Assign signs '+' and '–' alternately and find the minimum allocation from the cell having the negative sign. This allocation should be added to the allocation having the positive sign and subtracted from the allocation having the negative sign.

Step 6: The previous step yields a better solution by making one (or more) occupied cell as empty and one empty cell as occupied. For this new set of basic feasible allocations, repeat from Step (2) till an optimum basic feasible solution is obtained.

Example 5.7: Solve the following transportation problem.

		Destination				
		P	*Q*	*R*	*S*	*Supply*
Source	*A*	21	16	25	13	11
	B	17	18	14	23	13
	C	32	17	18	41	19
	Demand	6	10	12	15	43

Origin\Dest	P	Q	R	S	Supply	P_I	P_{II}	P_{III}	P_{IV}	P_V	P_{VI}
A	21	16	25	13 (11)	11	3	–	–	–	–	–
B	17 (6)	18	14 (3)	23 (4)	13	4	4	4	4 ←	–	–
C	32	17 (10)	18 (9)	48	19	1	1	1	1	1	17
Demand	6	10	12	15	43						
P_I	4	1	4	10↑							
P_{II}	15	1	4	18↑							
P_{III}	15↑	1	4	–							
P_{IV}	–	1	4	–							
P_V	–	17	18↑	–							
P_{VI}	–	17↑	–	–							

Solution: We first find the initial basic feasible solution by using VAM. Since $\Sigma a_i = \Sigma b_j$, the given TP is a balanced one. Therefore, there exists a feasible solution.

Finally, we have the initial basic feasible solution as given in the following table.

		Destination			
		P	Q	R	S
Source	A	21	16	23	13 (11)
	B	17 (6)	18	14 (3)	23 (4)
	C	32	17 (10)	18 (9)	41

From this table, we see that the number of non-negative independent allocations is $6 = m + n - 1 = 3 + 4 - 1$.

Hence, the solution is non-degenerate basic feasible.

Therefore, the initial transportation cost

$= 11 \times 13 + 3 \times 14 + 4 \times 23 + 6 \times 17 + 17 \times 10 + 18 \times 9 =$ ₹ 711

To find the optimal solution: We apply the MODI method in order to determine the optimum solution. We determine a set of numbers u_i and v_j for each row and column, with $u_i + v_j = C_{ij}$ for each occupied cell. To start with, we give $u_2 = 0$ as the second row has the maximum number of allocation.

Now, we find the sum u_i and v_j for each empty cell and enter at the bottom left corner of that cell.

$$C_{21} = u_2 + v_1 = 17 = 0 + v_1 = 17 \Rightarrow v_1 = 17$$

$$C_{23} = u_2 + v_3 = 14 = 0 + v_3 = 14 \Rightarrow v_3 = 14$$

$$C_{24} = u_2 + v_4 = 23 = 0 + v_4 = 23 \Rightarrow v_4 = 23$$

$$C_{14} = u_1 + v_4 = 13 = u_1 + 23 = 13 \Rightarrow u_1 = 10$$

$$C_{33} = u_3 + v_3 = 18 = u_3 + 14 = 18 \Rightarrow u_3 = 4$$

Next, we find the net evaluations $\Delta_{ji} = C_{ij} - (u_i + v_j)$ for each unoccupied cell and enter at the bottom right corner of that cell.

NOTES

Initial table

	P	Q	R	S	U_i
A	[21] 7 14	[16] 3 13	[23] 4 21	[13] (11)	$U_1 = 10$
B	[17] (6)	[18] 13 5	[14] (3)	[23] (4)	$U_2 = 0$
C	[32] 21 9	[17] (10)	[18] (9)	[41] 25 16	$U_3 = 4$
V_j	$V_1 = 17$	$V_2 = 13$	$V_3 = 14$	$V_4 = 23$	

Since all $\Delta_{ij} > 0$, the solution is optimal and unique. The optimum solution is given by,

$X_{14} = 11; X_{21} = 6; X_{23} = 3; X_{24} = 4; X_{32} = 10; X_{33} = 9$

The minimum transportation cost

$= 11 \times 13 + 17 \times 6 + 3 \times 14 + 4 \times 23 + 10 \times 17 + 9 \times 18 = ₹\ 711$

Example 5.8: Solve the following transportation problem starting with the initial solution obtained by VAM.

	D_1	D_2	D_3	D_4	Supply
O_1	[2] (3)	[2]	[2]	[1]	3
O_2	[10]	[8]	[5] (3)	[4] (4)	7
O_3	[7] (1)	[6] (3)	[6] (1)	[8]	5
Demand	4	3	4	4	15

Solution: Since $\Sigma a_i = \Sigma b_j$, the problem is a balanced TP. Therefore, there exists a feasible solution.

	D_1	D_2	D_3	D_4	Supply	P_I	P_{II}	P_{III}	P_{IV}	P_V	P_{VI}
O_1	[2] (3)	[2]	[2]	[1]	3	1	–	–	–	–	–
O_2	[10]	[8]	[5] (3)	[4] (4)	7	1	1	3 ←	–	–	–
O_3	[7] (1)	[6] (3)	[6] (1)	[8]	5	0	0	0	0	0	6 ←
Demand	4	3	4	4	15						
P_I	5↑	4	4	3							
P_{II}	3	2	1	4↑							
P_{III}	3	2	1	–							
P_{IV}	7↑	6	6	–							
P_V	–	6↑	6	–							
P_{VI}	–	–	6	–							

Finally, the initial basic feasible solution is given as follows:

	D_1	D_2	D_3	D_4	Supply
O_1	[2] (3)	[2]	[2]	[1]	3
O_2	[10]	[8]	[5] (3)	[4] (4)	7
O_3	[7] (1)	[6] (3)	[6] (1)	[8]	5
Demand	4	3	4	4	15

Since the number of occupied cells = 6 = $m + n - 1$ and are also independent, there exists a non-degenerate basic feasible solution.

The initial transportation cost

$= 3 \times 2 + 3 \times 5 + 4 \times 4 + 7 \times 1 + 6 \times 3 + 6 \times 1 = ₹\ 68$

To find the optimal solution, applying the MODI method, we determine a set of numbers u_i and v_j for each row and column, such that $u_i + v_j = C_{ij}$ for each occupied cell. Since the third now has the maximum number of allocations, we give number $u_3 = 0$. The remaining numbers can be obtained as follows:

$$C_{31} = u_3 + v_1 = 7 = 0 + v_1 = 7 \Rightarrow v_1 = 7$$
$$C_{32} = u_3 + v_2 = 6 = 0 + v_2 = 6 \Rightarrow v_2 = 6$$
$$C_{33} = u_3 + v_3 = 6 = 0 + v_3 = 6 \Rightarrow v_3 = 6$$
$$C_{23} = u_2 + v_3 = 5 = u_2 + 6 = 5 \Rightarrow u_2 = -1$$
$$C_{24} = u_2 + v_4 = 4 = -1 + v_4 = 4 \Rightarrow v_4 = 5$$
$$C_{11} = u_1 + v_1 = 2 = u_1 + 7 = 2 \Rightarrow u_1 = -5$$

We find the sum u_i and v_j for each empty cell and enter at the bottom left corner of the cell. Next, we find the net evaluation Δ_{ij} given by $\Delta_{ij} = C_{ij} - (u_i + v_j)$ for each empty cell and enter at the bottom right corner of the cell.

Initial table

	D_1	D_2	D_3	D_4	u_i
O_1	[2] ③	[2] 1 1	[2] 1 1	[1] 0 1	$u_1 = -5$
O_2	[10] 6 4	[8] 5 3	[5] ③	[4] ④	$u_2 = -1$
O_3	[7] ①	[6] ③	[6] ①	[8] 5 3	$u_3 = 0$
	$v_1 = 7$	$v_2 = 6$	$v_3 = 6$	$v_4 = 5$	

Since all $\Delta_{ij} > 0$, the solution is optimum and unique. The solution is given by,

$$X_{11} = 3;\ X_{23} = 3;\ X_{24} = 4,\ X_{31} = 1;\ X_{32} = 3;\ X_{33} = 1$$

The total transportation cost

$= 2 \times 3 + 3 \times 5 + 4 \times 4 + 7 \times 1 + 6 \times 3 + 6 \times 1 = ₹\ 68$

Check Your Progress

1. What is a transportation problem?
2. List the approaches used with transportation problems for determining the starting solution.
3. State the optimal solution to a transportation problem.
4. What is the purpose of the MODI method?
5. State the two conditions necessary for an alternate solution.

5.3 ASSIGNMENT PROBLEM

Suppose there are n jobs to be performed and n persons are available for doing these jobs. Assume that each person can do each job at a time, though with varying degrees of efficiency. Let c_{ij} be the cost if the ith person is assigned to the jth job. The problem is to find an assignment (which job should be assigned to which person, on a one to one basis) so that the total cost of performing all the jobs is minimum. Problems of this kind are known as assignment problems.

An assignment problem can be stated in the form of $n \times n$ cost matrix $[c_{ij}]$ of real numbers as given below:

Jobs

$$\textbf{Persons}\quad \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \\ i \\ \\ n \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & \cdots & j & \cdots & n \end{array} \\ \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1j} & \cdots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2j} & \cdots & c_{2n} \\ c_{31} & c_{32} & c_{33} & \cdots & c_{3j} & \cdots & c_{3n} \\ & & & & & & \\ c_{i1} & c_{i2} & c_{i3} & \cdots & c_{ij} & \cdots & c_{in} \\ \vdots & & & & & & \\ c_{n1} & c_{n2} & c_{n3} & \cdots & c_{nj} & \cdots & c_{nn} \end{bmatrix} \end{array}$$

5.3.1 Mathematical Formulation of an Assignment Problem

Mathematically, an assignment problem can be stated as,

$$\text{Minimize } z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}\, x_{ij}, \text{ where, } i = 1, 2, \ldots n, \text{ and } j = 1, 2, \ldots n$$

Subject to the restrictions,

$$x_{ij} = \begin{cases} 1, \text{ if the } i\text{th person is assigned } j\text{th job} \\ 0, \text{ if not} \end{cases}$$

$$\sum_{j=1}^{n} x_{ij} = 1 \text{ (one job is done by the } i\text{th person)}$$

$$\text{and } \sum_{j=1}^{n} x_{ij} = 1 \text{ (only one person should be assigned the } j\text{th job)}$$

where, x_{ij} denotes that the jth job is to be assigned to the ith person.

Difference between Transportation and Assignment Problems

Table 5.2 *Difference between Transportation and Assignment Problems*

Given below are the differences between transportation and assignment problem:

Transportation problem	*Assignment problem*
1. Number of sources and destinations need not be equal. Hence, the cost matrix is not necessarily a square matrix.	Since assignment is done on a one to one basis, the number of sources and destinations are equal. Hence, the cost matrix must be a square matrix.
2. x_{ij}, the quantity to be transported from ith origin to jth destination can take any possible positive value, and it satisfies the rim requirements.	x_{ij}, the jth job is to be assigned to the ith person and can take either the value 1 or zero.

3. The capacity and the requirement value is equal to a_i and b_j for the ith source and jth destination ($i = 1, 2, ...m;\ j = 1, 2, ...\ n$).	The capacity and the requirement value is exactly one, i.e., for each source of each destination, the capacity and the requirement value is exactly one.
4. The problem is unbalanced if the total supply and total demand are not equal.	The problem is unbalanced if the cost matrix is not a square matrix.

5.3.2 Hungarian Method Algorithm

Solution of an assignment problem can be arrived at, by using the **Hungarian method**. The steps involved in this method are as follows.

Step 1 Prepare a cost matrix. If the cost matrix is not a square matrix then add a dummy row (column) with zero cost element.

Step 2 Subtract the minimum element in each row from all the elements of the respective rows.

Step 3 Further, modify the resulting matrix by subtracting the minimum element of each column from all the elements of the respective columns. Thus, obtain the modified matrix.

Step 4 Then, draw the minimum number of horizontal and vertical lines to cover all zeros in the resulting matrix. Let the minimum number of lines be N. Now there are two possible cases.

Case I If $N = n$, where n is the order of matrix, then an optimal assignment can be made. So make the assignment to get the required solution.

Case II If $N < n$, then proceed to step 5.

Step 5 Determine the smallest uncovered element in the matrix (element not covered by N lines). Subtract this minimum element from all uncovered elements and add the same element at the intersection of horizontal and vertical lines. Thus, the second modified matrix is obtained.

Step 6 Repeat steps 3 and 4 until we get the case (i) of Step 4.

Step 7 (To make zero assignment) Examine the rows successively until a row-wise exactly single zero is found. Circle (O) this zero to make the assignment. Then mark a cross (×) over all zeros if lying in the column of the circled zero, showing that they cannot be considered for future assignment. Continue in this manner until all the zeros have been examined. Repeat the same procedure for columns also.

Step 8 Repeat step 6 successively until one of the following situation arises—

(*i*) If no unmarked zero is left, then the process ends or

(*ii*) If there lie more than one unmarked zero in any column or row, circle one of the unmarked zeros arbitrarily and mark a cross in the cells of remaining zeros in its row or column. Repeat the process until no unmarked zero is left in the matrix.

Step 9 Thus, exactly one marked circled zero in each row and each column of the matrix is obtained. The assignment corresponding to these marked circled zeros will give the optimal assignment.

Example 5.9: Using the following cost matrix, determine (*a*) optimal job assignment (*b*) the cost of assignments.

$$\textbf{Mechanic}\quad \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \overset{\textbf{Job}}{\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \left[\begin{array}{ccccc} 10 & 3 & 3 & 2 & 8 \\ 9 & 7 & 8 & 2 & 7 \\ 7 & 5 & 6 & 2 & 4 \\ 3 & 5 & 8 & 2 & 4 \\ 9 & 10 & 9 & 6 & 10 \end{array}\right] \end{array}}$$

Solution: Select the smallest element in each row and subtract this smallest element from all the elements in its row.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \left[\begin{array}{ccccc} 8 & 1 & 1 & 0 & 6 \\ 7 & 5 & 6 & 0 & 5 \\ 5 & 3 & 4 & 0 & 2 \\ 1 & 3 & 6 & 0 & 2 \\ 3 & 4 & 3 & 0 & 4 \end{array}\right] \end{array}$$

Select the minimum element from each column and subtract from all other elements in its column. With this we get the first modified matrix.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \left[\begin{array}{ccccc} 7 & 0 & 0 & 0 & 4 \\ 6 & 4 & 5 & 0 & 3 \\ 4 & 2 & 3 & 0 & 0 \\ 0 & 2 & 5 & 0 & 0 \\ 2 & 3 & 2 & 0 & 2 \end{array}\right] \end{array}$$

In this modified matrix we draw the minimum number of lines to cover all zeros (horizontal or vertical).

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \left[\begin{array}{ccccc} 7 & 0 & 0 & 0 & 4 \\ 6 & 4 & 5 & 0 & 3 \\ 4 & 2 & 3 & 0 & 0 \\ 0 & 2 & 5 & 0 & 0 \\ 2 & 3 & 2 & 0 & 2 \end{array}\right] \end{array}$$

Number of lines drawn to cover all zeros is $4 = N$.

The order of matrix is $n = 5$

Hence, $N < n$.

Now we get the second modified matrix by subtracting the smallest uncovered element from the remaining uncovered elements and adding it to the element at the point of intersection of lines.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \left[\begin{array}{ccccc} 9 & 0 & \not{0} & 2 & 6 \\ 6 & 2 & 3 & 0 & 3 \\ 4 & \not{0} & 1 & \not{0} & 0 \\ 0 & \not{0} & 3 & \not{0} & \not{0} \\ 2 & 1 & 0 & \not{0} & 2 \end{array}\right] \end{array}$$

NOTES

Number of lines drawn to cover all zeros = $N = 5$

The order of matrix is $n = 5$.

Hence $N = n$. Now we determine the optimum assignment.

Assignment

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \left[\begin{array}{ccccc} 9 & 0 & \cancel{0} & 2 & 6 \\ 6 & 2 & 3 & 0 & 3 \\ 4 & \cancel{0} & 1 & \cancel{0} & 0 \\ 0 & \cancel{0} & 3 & \cancel{0} & \cancel{0} \\ 2 & 1 & 0 & \cancel{0} & 2 \end{array}\right] \end{array}$$

First row contains more than one zero. So proceed to the 2nd row. It has exactly one zero. The corresponding cell is $(B, 4)$. Circle this zero thus, making an assignment. Mark (×) for all other zeros in its column. Showing that they cannot be used for making other assignments. Now row 5 has a single zero in the cell $(E, 3)$. Make an assignment in this cell and cross the 2nd zero in the 3rd column.

Now row 1 has a single zero in the column 2, i.e., in the cell $(A, 2)$. Make an assignment in this cell and cross the other zeros in the 2nd column. This leads to a single zero in column 1 of the cell $(D, 1)$, make an assignment in this cell and cross the other zeros in the 4th row. Finally, we have a single zero left in the 3rd row, making an assignment in the cell $(C, 5)$. Thus, we have the following assignment.

Optimal assignment and optimum cost of assignment.

Job	*Mechanic*	*Cost*
1	*D*	3
2	*A*	3
3	*E*	9
4	*B*	2
5	*C*	4
		₹ 21

Therefore, $1 \to D$, $2 \to A$, $3 \to E$, $4 \to B$, $5 \to C$, with minimum cost equal to ₹ 21.

Example 5.10: A company has 5 jobs to be done on five machines. Any job can be done on any machine. The cost of doing the jobs on different machines are given below. Assign the jobs for different machines so as to minimize the total cost.

	Machines				
Jobs	*A*	*B*	*C*	*D*	*E*
1	13	8	16	18	19
2	9	15	24	9	12
3	12	9	4	4	4
4	6	12	10	8	13
5	15	17	18	12	20

Solution: We form the first modified matrix by subtracting the minimum element from all the elements in the respective row, and the same with respective columns.

Step 1

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 5 & 0 & 8 & 10 & 11 \\ 0 & 6 & 15 & 0 & 3 \\ 8 & 5 & 0 & 0 & 0 \\ 0 & 6 & 4 & 2 & 7 \\ 3 & 5 & 6 & 0 & 8 \end{bmatrix} \end{array}$$

Since each column has the minimum element 0, we have the first modified matrix. Now we draw the minimum number of lines to cover all zeros.

Step 2

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 5 & 0 & 8 & 10 & 11 \\ 0 & 6 & 15 & 0 & 3 \\ 8 & 5 & 0 & 0 & 0 \\ 0 & 6 & 4 & 2 & 7 \\ 3 & 5 & 6 & 0 & 8 \end{bmatrix} \end{array}$$

Step 3

$$\begin{bmatrix} 5 & 0 & 5 & 10 & 8 \\ 0 & 6 & 12 & 0 & 0 \\ 11 & 8 & 0 & 3 & 0 \\ 0 & 6 & 1 & 2 & 4 \\ 3 & 5 & 3 & 0 & 5 \end{bmatrix}$$

Here, the smallest uncovered element is '3'. So '3' is added to all the junction points i.e., the prints of intersection of lines.

Number of lines drawn to cover zero is $N = 4 <$ the order of matrix $n = 5$.

We find the second modified matrix by subtracting the smallest uncovered element (3) from all the uncovered elements and adding to the element that is the point of intersection of lines.

Step 4

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 5 & 0 & 5 & 10 & 8 \\ 0 & 6 & 12 & 0 & 0 \\ 11 & 8 & 0 & 3 & 0 \\ 0 & 6 & 1 & 2 & 4 \\ 3 & 5 & 3 & 0 & 5 \end{bmatrix} \end{array}$$

Number of lines drawn to cover all zeros = 5,

which is the order of matrix. Hence, we can form an assignment.

Assignment

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 5 & 0 & 5 & 10 & 8 \\ \cancel{0} & 6 & 12 & \cancel{0} & 0 \\ 11 & 8 & 0 & 3 & \cancel{0} \\ 0 & 6 & 1 & 2 & 4 \\ 3 & 5 & 3 & 0 & 5 \end{bmatrix} \end{array}$$

All the five jobs have been assigned to 5 different machines.

Here the optimal assignment is,

Job	***Machine***
1	B
2	E
3	C
4	A
5	D

Minimum (Total cost) = 8 + 12 + 4 + 6 + 12 = ₹ 42.

Example 5.11: Four different jobs can be done on four different machines and the take-down time costs are prohibitively high for change overs. The matrix below gives the cost in rupees for producing job i on machine j.

	Machines			
Jobs	M_1	M_2	M_3	M_4
J_1	5	7	11	6
J_2	8	5	9	6
J_3	4	7	10	7
J_4	10	4	8	3

How should the jobs be assigned to the various machines so that the total cost is minimized.

Solution: We form a first modified matrix by subtracting the least element in the respective rows and respective columns.

$$\begin{array}{c} \\ J_1 \\ J_2 \\ J_3 \\ J_4 \end{array} \begin{array}{c} \begin{matrix} M_1 & M_2 & M_3 & M_4 \end{matrix} \\ \begin{bmatrix} 0 & 2 & 6 & 1 \\ 3 & 0 & 4 & 1 \\ 0 & 3 & 6 & 3 \\ 7 & 1 & 5 & 0 \end{bmatrix} \end{array}$$

Since the third column has no zero element, we subtract the smallest element 4 from all the elements.

$$\begin{array}{c} \\ J_1 \\ J_2 \\ J_3 \\ J_4 \end{array} \begin{array}{c} \begin{matrix} M_1 & M_2 & M_3 & M_4 \end{matrix} \\ \begin{bmatrix} 0 & 2 & 2 & 1 \\ 3 & 0 & 0 & 1 \\ 0 & 3 & 2 & 3 \\ 7 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

Now we draw minimum number of lines to cover all zeros.

$$\begin{array}{c} \\ J_1 \\ J_2 \\ J_3 \\ J_4 \end{array} \begin{array}{c} \begin{matrix} M_1 & M_2 & M_3 & M_4 \end{matrix} \\ \begin{bmatrix} 0 & 2 & 2 & 1 \\ 3 & 0 & 0 & 1 \\ 0 & 3 & 2 & 3 \\ 7 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

Number of lines drawn to cover all zeros = 3, which is less than the order of matrix = 4.

Hence, we form the 2nd modified matrix, by subtracting the smallest uncovered element from all the uncovered elements and adding to the element that is at the point of intersection of lines.

$$\begin{array}{c} \\ J_1 \\ J_2 \\ J_3 \\ J_4 \end{array}\begin{array}{c} \begin{array}{cccc} M_1 & M_2 & M_3 & M_4 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 1 \\ 4 & 0 & 0 & 2 \\ 0 & 2 & 1 & 3 \\ 7 & 0 & 0 & 0 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ J_1 \\ J_2 \\ J_3 \\ J_4 \end{array}\begin{array}{c} \begin{array}{cccc} M_1 & M_2 & M_3 & M_4 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 \\ 8 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$N = 3 < n = 4 \qquad\qquad N = 4 = n = 4$$

Hence, we can make an assignment.

$$\begin{array}{c} \\ J_1 \\ J_2 \\ J_3 \\ J_4 \end{array}\begin{array}{c} \begin{array}{cccc} M_1 & M_2 & M_3 & M_4 \end{array} \\ \begin{bmatrix} \otimes & \otimes & \otimes & 0 \\ 5 & 0 & \otimes & 2 \\ 0 & 1 & \otimes & 2 \\ 8 & \otimes & 0 & \otimes \end{bmatrix} \end{array}$$

Since no rows and no columns have single zero, we have a different assignment (Multiple solution).

Optimal assignment

Job	***Machine***
J_1	M_4
J_2	M_2
J_3	M_1
J_4	M_3

Minimum (Total cost)

$$6 + 5 + 4 + 8 = ₹\ 23$$

Alternate Solution

$$J_1 \to M_1; \quad J_2 \to M_2; \quad J_3 \to M_3; \quad J_4 \to M_4.$$

Minimum (Total cost)

$$5 + 5 + 10 + 3 = ₹\ 23.$$

Example 5.12: Solve the following assignment problem in order to minimize the total cost. The cost matrix given below gives the assignment cost when different operators are assigned to various machines.

Operators

Machines	*I*	*II*	*III*	*IV*	*V*
A	30	25	33	35	36
B	23	29	38	23	26
C	30	27	22	22	22
D	25	31	29	27	32
E	27	29	30	24	32

Solution: We form the first modified matrix by subtracting the least element from all the elements in the respective rows and then in the respective columns.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} I & II & III & IV & V \end{array} \\ \begin{bmatrix} 5 & 0 & 8 & 10 & 11 \\ 0 & 6 & 15 & 0 & 3 \\ 8 & 5 & 0 & 0 & 0 \\ 0 & 6 & 4 & 2 & 7 \\ 3 & 5 & 6 & 0 & 8 \end{bmatrix} \end{array}$$

Since each column has the minimum element 0, the first modified matrix is obtained. We draw the minimum number of lines to cover all zeros.

The number of lines drawn to cover all zeros = 4 < the order of matrix = 5. Hence, we form the second modified matrix by subtracting the smallest uncovered element from the remaining uncovered elements and adding to the element that is at the point of intersection of lines.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} I & II & III & IV & V \end{array} \\ \begin{bmatrix} 5 & 0 & 5 & 10 & 8 \\ 0 & 6 & 12 & 0 & 0 \\ 11 & 8 & 0 & 3 & 0 \\ 0 & 6 & 1 & 2 & 4 \\ 3 & 5 & 3 & 0 & 5 \end{bmatrix} \end{array}$$

$N = 5$, i.e., the number of lines drawn to cover all zeros = order of matrix. Hence, we can make an assignment.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} I & II & III & IV & V \end{array} \\ \begin{bmatrix} 5 & 0 & 5 & 10 & 8 \\ \cancel{0} & 6 & 12 & \cancel{0} & 0 \\ 11 & 8 & 0 & 3 & \cancel{0} \\ 0 & 6 & 1 & 2 & 4 \\ 3 & 5 & 3 & 0 & 5 \end{bmatrix} \end{array}$$

The optimum assignment is

Operators	***Machines***
I	D
II	A
III	C
IV	E
V	B

The optimum cost is given by

$$25 + 25 + 22 + 24 + 26 = ₹\ 122.$$

Unbalanced Assignment problem

Any assignment problem is said to be unbalanced if the cost matrix is not a square matrix, i.e., the number of rows and columns are not equal. To make it balanced, we add a dummy row or dummy column with all the entries as zero.

NOTES

Example 5.13: There are four jobs to be assigned to five machines. Only one job can be assigned to one machine. The amount of time in hours required for the jobs per machine are given in the following matrix.

	Machines				
Jobs	*A*	*B*	*C*	*D*	*E*
1	4	3	6	2	7
2	10	12	11	14	16
3	4	3	2	1	5
4	8	7	6	9	6

Find an optimum assignment of jobs to the machines to minimize the total processing time and also find out for which machine no job is assigned. What is the total processing time to complete all the jobs?

Solution Since the cost matrix is not a square matrix, the problem is unbalanced. We add a dummy job 5 with corresponding entries zero.

Modified Matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 4 & 3 & 6 & 2 & 7 \\ 10 & 12 & 11 & 14 & 16 \\ 4 & 3 & 2 & 1 & 5 \\ 8 & 7 & 6 & 9 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

We subtract the smallest element from all the elements in the respective rows.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 2 & 1 & 4 & 0 & 5 \\ 0 & 2 & 1 & 4 & 6 \\ 3 & 2 & 1 & 0 & 4 \\ 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Since each column has minimum element as zero, we draw minimum number of lines to cover all zeros.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 2 & 1 & 4 & 0 & 5 \\ 0 & 2 & 1 & 4 & 6 \\ 3 & 2 & 1 & 0 & 4 \\ 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

The number of lines to cover all zeros = 4 < the order of matrix. We form the 2nd modified matrix by subtracting the smallest uncovered element from the remaining uncovered elements and adding to the element at the point of intersection of lines.

NOTES

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 2 & 0 & 3 & \cancel{0} & 4 \\ 0 & 1 & \cancel{0} & 4 & 5 \\ 3 & 1 & 0 & \cancel{0} & 3 \\ 3 & 1 & \cancel{0} & 4 & 0 \\ 1 & \cancel{0} & \cancel{0} & 1 & \cancel{0} \end{bmatrix} \end{array}$$

Here the number of lines drawn to cover all zeros = 5 = Order of matrix. Therefore, we can make the assignment

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 2 & 0 & 3 & 0 & 4 \\ 0 & 1 & 0 & 4 & 5 \\ 3 & 1 & 0 & 0 & 3 \\ 3 & 1 & 0 & 4 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

Optimum assignment

1	B	3
2	A	10
3	C	2
4	D	6

For machine *D*, no job is assigned.

Optimum (minimum) cost = 3 + 10 + 1 + 6 = ₹ 20.

Example 5.14: A company has 4 machines to do 3 jobs. Each job can be assigned to only one machine. The cost of each job on each machine is given below. Determine the job assignments that will minimize the total cost.

Machine

$$\textbf{Job}\ \begin{array}{c} \\ A \\ B \\ C \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 18 & 24 & 28 & 32 \\ 8 & 13 & 17 & 18 \\ 10 & 15 & 19 & 22 \end{bmatrix} \end{array}$$

Solution: Since the cost matrix is not a square matrix, we add a dummy row *D* with all the elements 0.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 18 & 24 & 28 & 32 \\ 8 & 13 & 17 & 18 \\ 10 & 15 & 19 & 22 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Subtract the minimum element in each row from all the elements in its row.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 6 & 10 & 14 \\ 0 & 5 & 9 & 10 \\ 0 & 5 & 9 & 12 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Since each column has a minimum element 0, we draw minimum number of lines to cover all zeros.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 6 & 10 & 14 \\ 0 & 5 & 9 & 10 \\ 0 & 5 & 9 & 12 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$\therefore$ The number of lines drawn to cover all zeros = 2 < the order of matrix, we form a second modified matrix.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 5 & 9 \\ 0 & 0 & 4 & 5 \\ 0 & 0 & 4 & 7 \\ 5 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Here, $N = 3 < n = 4$.

Again we subtract the smallest uncovered element from all the uncovered elements and add to the element at the point of intersection

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 3 \\ 9 & 4 & 0 & 0 \end{bmatrix} \end{array}$$

Here, $N = 4 = n$. Hence, we make an assignment.

Assignment

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 4 \\ \not 0 & 0 & \not 0 & 1 \\ \not 0 & \not 0 & 0 & 3 \\ 9 & 4 & \not 0 & 0 \end{bmatrix} \end{array} \begin{array}{l} A \to W \\ D \to Z \\ B \to X \\ C \to Y \end{array} \quad \text{or} \quad \begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 4 \\ \not 0 & \not 0 & 0 & 1 \\ \not 0 & 0 & \not 0 & 3 \\ 9 & 4 & \not 0 & 0 \end{bmatrix} \end{array} \begin{array}{l} A \to W \\ D \to Z \\ B \to Y \\ C \to X \end{array}$$

Since D is a dummy job, machine Z is assigned no job.

Therefore, optimum cost = 18 + 13 + 19 = ₹ 50.

Maximization in Assignment problem

In this, the objective is to maximize the profit. To solve this, we first convert the given profit matrix into the loss matrix by subtracting all the elements from the highest element. For this converted loss matrix we apply the steps in Hungarian method to get the optimum assignment.

Example 5.15: The owner of a small machine shop has four mechanics available to assign jobs for the day. Five jobs are offered with expected profit for each mechanic on each jobs, which are as follows:

Job

Mechanic		A	B	C	D	E
	1	62	78	50	111	82
	2	71	84	61	73	59
	3	87	92	111	71	81
	4	48	64	87	77	80

By using the assignment method, find the assignment of mechanics to the job that will result in maximum profit. Which job should be declined ?

Solution: The given profit matrix is not a square matrix as the number of jobs is not equal to the number of mechanics. Hence, we introduce a dummy mechanic 5 with all the elements 0.

Job

Mechanic		A	B	C	D	E
	1	62	78	50	111	82
	2	71	84	61	73	59
	3	87	92	111	71	81
	4	48	64	87	77	80
	5	0	0	0	0	0

Now we convert this profit matrix into loss matrix by subtracting all the elements from the highest element 111.

Loss Matrix

	A	B	C	D	E
1	49	33	61	0	29
2	40	27	50	38	52
3	24	19	0	40	30
4	63	47	24	34	31
5	111	111	111	111	111

We subtract the smallest element from all the elements in the respective rows.

	A	B	C	D	E
1	49	33	61	0	29
2	13	0	23	11	25
3	24	19	0	40	30
4	39	23	0	10	7
5	0	0	0	0	0

Since each column has minimum element as zero, we draw minimum number of lines to cover all zeros.

NOTES

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 49 & 33 & 61 & 0 & 29 \\ 13 & 0 & 23 & 11 & 25 \\ 24 & 19 & 0 & 40 & 30 \\ 39 & 23 & 0 & 10 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Here the number of lines drawn to cover all zeros = $N = 4$, is less than the order of matrix.

We form the 2nd modified matrix by subtracting the smallest uncovered element from the remaining uncovered elements and adding to the element that is at the point of intersection of lines.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 49 & 40 & 68 & 0 & 29 \\ 6 & 0 & 23 & 5 & 18 \\ 17 & 19 & 0 & 33 & 23 \\ 32 & 23 & 0 & 3 & 0 \\ 0 & 7 & 7 & 0 & 0 \end{bmatrix} \end{array}$$

Here, $N = 5 = n$ (the order of matrix).

We make the assignment.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 49 & 40 & 68 & 0 & 29 \\ 6 & 0 & 23 & 5 & 18 \\ 17 & 19 & 0 & 33 & 23 \\ 32 & 23 & \not{0} & 3 & 0 \\ 0 & 7 & 7 & \not{0} & \not{0} \end{bmatrix} \end{array}$$

The optimum assignment is

Job	***Mechanic***
A	5
B	2
C	3
D	1
E	4

Since the 5th mechanic is a dummy, job *A* is assigned to the 5th mechanic, this job is declined.

The maximum profit is given by, 84 + 111 + 111 + 80 = ₹ 386.

Example 5.16: A marketing manager has 5 salesmen and there are 5 sales districts. Considering the capabilities of the salesmen and the nature of districts, the estimates made by the marketing manager for the sales per month (in 1,000 rupees) for each salesman in each district would be as follows.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 32 & 38 & 40 & 28 & 40 \\ 40 & 24 & 28 & 21 & 36 \\ 41 & 27 & 33 & 30 & 37 \\ 22 & 38 & 41 & 36 & 36 \\ 29 & 33 & 40 & 35 & 39 \end{bmatrix} \end{array}$$

Find the assignment of salesmen to the districts that will result in the maximum sales.

Solution: We are given the profit matrix. To maximize the profit, first we convert it into a loss matrix, which can be minimized. To convert it into loss matrix, we subtract all the elements from the highest element 41. Subtract the smallest element from all the elements in the respective rows and columns, to get the first modified matrix.

Loss Matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 9 & 3 & 1 & 13 & 1 \\ 1 & 17 & 13 & 20 & 5 \\ 0 & 14 & 8 & 11 & 4 \\ 19 & 3 & 0 & 5 & 5 \\ 12 & 8 & 1 & 6 & 2 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 8 & 2 & 0 & 12 & 0 \\ 0 & 16 & 12 & 19 & 4 \\ 0 & 14 & 8 & 11 & 4 \\ 19 & 3 & 0 & 5 & 5 \\ 11 & 7 & 0 & 5 & 1 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 8 & 0 & 0 & 7 & 0 \\ 0 & 14 & 12 & 14 & 4 \\ 0 & 12 & 8 & 6 & 4 \\ 19 & 1 & 0 & 0 & 5 \\ 11 & 5 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

We now draw minimum number of lines to cover all zeros.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 8 & 0 & 0 & 7 & 0 \\ 0 & 14 & 12 & 4 & 4 \\ 0 & 12 & 8 & 6 & 4 \\ 9 & 1 & 0 & 0 & 5 \\ 11 & 5 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$N = 4 < n = 5$

NOTES

We subtract the smallest uncovered element from the remaining uncovered elements and add to the elements at the point of intersection of lines, to get the second modified matrix.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 9 & 0 & 1 & 8 & 0 \\ 0 & 13 & 12 & 4 & 3 \\ 0 & 11 & 8 & 6 & 3 \\ 9 & 0 & 0 & 0 & 4 \\ 11 & 4 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Again, $N = 4$ $n = 5$. Repeat the above step.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 12 & 0 & 1 & 8 & 0 \\ 0 & 10 & 9 & 1 & 0 \\ 0 & 8 & 5 & 3 & 0 \\ 12 & 0 & 0 & 0 & 4 \\ 14 & 4 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$N = 5 = n = 5$. Hence we make the assignment.

Assignment

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 12 & 0 & 1 & 8 & \times \\ 0 & 10 & 9 & 1 & \times \\ \times & 8 & 5 & 3 & 0 \\ 12 & \times & 0 & \times & 4 \\ 14 & 4 & \times & 0 & \times \end{bmatrix} \end{array}$$

Since no row or column has single zero, we get a multiple solution.

(*i*) The optimum assignment is:

$1 \rightarrow B, 2 \rightarrow A, 3 \rightarrow E, 4 \rightarrow C, 5 \rightarrow D.$

With maximum profit $(38 + 40 + 37 + 41 + 35) =$ ₹ 191

(*ii*) The optimum assignment is:

$1 \rightarrow B, 2 \rightarrow A, 3 \rightarrow E, 4 \rightarrow D, 5 \rightarrow C.$

Maximum profit $(38 + 40 + 37 + 36 + 40) =$ ₹ 191

5.3.3 Routing Problem: The Travelling Salesman Problem

A network routing problem consists of finding an optimum route between two or more nodes in relation to total time, cost, or distance. Various constraints may exist, such as a prohibition on returning to a node already visited or a stipulation of passing through every node only once. Network routing problems commonly arise in communication and transportation systems. Delays that occur at the nodes (*e.g.*, railroad classification yards or telephone switchboards) may be a function of the loads placed on them and their capacities. Breakdowns may occur in either links or nodes. Much studied is the "traveling salesman problem," which consists of starting a route from a designated node that goes through each node (*e.g.*, city) only once and returns to the origin in the least time, cost, or distance. This problem arises in selecting an order for processing a set of production jobs when the cost of setting up each job depends on which job has preceded it.

Let us discuss the 'Traveling Salesman Problem'.

Assuming a salesman has to visit n cities. He wishes to start from a particular city, visit each city once and then return to his starting point. His objective is to select the sequence in which the cities are visited in such a way that his total travelling time is minimized.

To visit 2 cities (A and B, there is no choice. To visit 3 cities we have 2 possible routes. For 4 cities we have 3 possible routes. In general, to visit n cities there are $(n-1)$ possible routes.

Mathematical Formulation

Let C_{ij} be the distance or time or cost of going from city i to city j. Let the decision variable X_{ij} be 1, if the salesman travels from city i to city j, otherwise let it be 0.

The objective is to minimize the travelling time.

$$Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}\, x_{ij}$$

Subject to the constraints,

$$\sum_{j=1}^{n} X_{ij} = 1,\ i = 2 \dots n.$$

$$\sum_{i=1}^{n} X_{ij} = 1,\ j = 2 \dots n.$$

and subject to the additional constraint that X_{ij} is so chosen that, no city is visited twice before all the cities are visited.

In particular, going from i directly to i is not permitted. This means $C_{ij} = \infty$, when $i = j$.

In the travelling salesman problem we cannot choose the element along the diagonal and this can be avoided by filling the diagonal with infinitely large elements.

The travelling salesman problem is very similar to the assignment problem except that in the former case, there is an additional restriction, that X_{ij} is so chosen that no city is visited twice before the tour of all the cities is completed.

Example 5.17: A travelling salesman has to visit 5 cities. He wishes to start from a particular city, visit each city once and then return to his starting point. Cost of going from one city to another is shown below. You are required to find the least cost route.

To City

From City

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 4 & 10 & 14 & 2 \\ 12 & \infty & 6 & 10 & 4 \\ 16 & 14 & \infty & 8 & 14 \\ 24 & 8 & 12 & \infty & 10 \\ 2 & 6 & 4 & 16 & \infty \end{bmatrix} \end{array}$$

Solution: Treat the problem as an assignment problem and solve it using the same procedures. If the optimal solution of the assignment problem satisfies the additional constraint, then it is also an optimal solution of the given travelling salesman problem. If the solution to the assignment problem does not satisfy the additional restriction, then after solving the problem by assignment technique, we use the method of enumeration.

First we solve this problem as an assignment problem.

Subtract the minimum element in each row from all the elements in its row.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 2 & 8 & 12 & 0 \\ 8 & \infty & 2 & 6 & 0 \\ 8 & 6 & \infty & 0 & 6 \\ 16 & 0 & 4 & \infty & 2 \\ 0 & 4 & 2 & 14 & \infty \end{bmatrix} \end{array}$$

Subtract the minimum element in each column from all the elements in its column.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 2 & 8 & 12 & 0 \\ 8 & \infty & 2 & 6 & 0 \\ 8 & 6 & \infty & 0 & 6 \\ 16 & 0 & 4 & \infty & 2 \\ 0 & 4 & 2 & 14 & \infty \end{bmatrix} \end{array}$$

We have the first modified matrix. Draw minimum number of lines to cover all zeros.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 2 & 6 & 12 & 0 \\ 8 & \infty & 2 & 6 & 0 \\ 8 & 8 & \infty & 0 & 6 \\ 16 & 0 & 2 & \infty & 2 \\ 0 & 4 & 0 & 14 & \infty \end{bmatrix} \end{array}$$

$N = 4 < n = 5$. Subtract the smallest uncovered element from all the uncovered elements and add to the element that is at the point of intersection of lines. Hence, we get the 2nd modified matrix.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 0 & 6 & 12 & 0 \\ 8 & \infty & 0 & 6 & 0 \\ 8 & 4 & \infty & 0 & 6 \\ 18 & 0 & 4 & \infty & 4 \\ 0 & 4 & 0 & 14 & \infty \end{bmatrix} \end{array}$$

$N = 5 = n = 5$. We make an assignment.

Assignment

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & \boxtimes & 6 & 12 & 0 \\ 8 & \infty & 0 & 6 & \boxtimes \\ 8 & 4 & \infty & 0 & 6 \\ 18 & 0 & 4 & \infty & 4 \\ 0 & 4 & \boxtimes & 14 & \infty \end{bmatrix} \end{array} \quad \begin{array}{l} A \to E \\ B \to C \\ C \to D \\ D \to B \\ E \to A \\ A \to E \to A \end{array}$$

The salesman should go from A to E and then come back to A without covering B, C, D. But this is contradicting the constraint that no city is visited twice before all the cities are visited.

Hence, we obtain the next best solution by bringing the next minimum non-zero element namely 4.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 0 & 6 & 12 & 0 \\ 8 & \infty & 0 & 6 & \boxtimes \\ 8 & 4 & \infty & 0 & 6 \\ 18 & \boxtimes & 4 & \infty & 4 \\ 0 & 4 & \boxtimes & 14 & \infty \end{bmatrix} \end{array}$$

$A \to B,\ B \to C,\ C \to D,\ D \to E,\ E \to A$

Since all the cities have been visited and no city is visited twice before completing the tour of all the cities, we have an optimal solution for the travelling salesman.

The least cost route is $A \to B \to C \to D \to E \to A$.

Total cost = 4 + 6 + 8 + 10 + 2 = ₹ 30.

Example 5.18: A machine operator processes five types of items on his machine each week and must choose a sequence for them. The set-up cost per change depends on the items presently on the machine and the set-up to be made, according to the following table.

NOTES

To Item

$$\text{From Item}\quad \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 4 & 7 & 3 & 4 \\ 4 & \infty & 6 & 3 & 4 \\ 7 & 6 & \infty & 7 & 5 \\ 3 & 3 & 7 & \infty & 7 \\ 4 & 4 & 5 & 7 & \infty \end{bmatrix} \end{array}$$

If he processes each type of item only once in each week, how should he sequence the items on his machine in order to minimize the total set-up cost?

Solution: Reduce the cost matrix and make assignments in rows and columns having single row.

Modify the matrix by subtracting the least element from all the elements in its row and also in its column.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 1 & 4 & 0 & 1 \\ 1 & \infty & 3 & 0 & 1 \\ 2 & 1 & \infty & 2 & 0 \\ 0 & 0 & 4 & \infty & 4 \\ 0 & 0 & 1 & 3 & \infty \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 1 & 3 & 0 & 1 \\ 1 & \infty & 2 & 0 & 1 \\ 2 & 1 & \infty & 2 & 0 \\ 0 & 0 & 3 & \infty & 4 \\ 0 & 0 & 0 & 3 & \infty \end{bmatrix} \end{array}$$

Here, $N = 4$ $n = 5$, i.e., $N < n$.

Subtract the smallest uncovered element from all the uncovered elements and add to the element that is at the point of intersection of lines and get the reduced 2nd modified matrix.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 0 & 2 & 0 & 1 \\ 0 & \infty & 1 & 0 & 1 \\ 1 & 0 & \infty & 1 & 0 \\ 0 & 0 & 3 & \infty & 5 \\ 0 & 0 & 0 & 4 & \infty \end{bmatrix} \end{array}$$

$N = 5 = n = 5$. We make the assignment.

NOTES

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \left[\begin{array}{ccccc} \infty & \not{0} & 2 & \not{0} & 1 \\ \not{0} & \infty & 1 & 0 & 1 \\ 1 & 0 & \infty & 1 & \not{0} \\ 0 & \not{0} & 3 & \infty & 5 \\ \not{0} & \not{0} & 0 & 4 & \infty \end{array}\right] \end{array}\begin{array}{l} \\ A \to B \\ B \to D \\ C \to E \\ D \to A \\ E \to C \end{array}$$

We get the solution $A \to B \to D \to A$.

This schedule does not provide the required solution as each item is not processed only once in a week.

Hence, we make a better solution by considering the next smallest non-zero element by considering 1.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \left[\begin{array}{ccccc} \infty & \not{0} & 2 & \not{0} & 1 \\ \not{0} & \infty & \not{1} & 0 & \not{1} \\ 1 & 0 & \infty & 1 & \not{0} \\ 0 & \not{0} & 3 & \infty & 5 \\ \not{0} & \not{0} & 0 & 4 & \infty \end{array}\right] \end{array}$$

$A \to E$, $E \to C$, $C \to B$, $B \to D$, $D \to A$,

i.e., $A \to E \to C \to B \to D \to A$.

The total set-up cost comes to ₹ 21.

Check Your Progress

6. State any one difference between transportation and assignment problems.
7. What is an unbalanced assignment problem?
8. What do you mean by routing problem?

5.4 INTEGER PROGRAMMING

In LPP, all the decision variables were allowed to take any non-negative real values as it is quite possible and appropriate to have fractional values in many situations. There are several frequently occurring circumstances in business and industry that lead to planning models involving integer-valued variables. For example, in production, manufacturing is frequently scheduled in terms of batches, lots or runs. In allocation of goods, a shipment must involve a discrete number of trucks or aircrafts. In such cases the fractional values of variables like 13/3 may be meaningless in the context of the actual decision problem.

Applications of Integer programming

Integer programming is applied in business and industry. All assignment and transportation problems are integer programming problems, as in the assignment and travelling salesmen problem, all the decision variables are either zero or one.

i.e., $x_{ij} = 0$ or 1

Other examples are capital budgeting and production scheduling problems. In fact, any situation involving decisions of the type 'either to do a job or not' can be viewed as an IPP In all such situations,

$x_{ij} = 1$, if the jth activity is performed,

0 if the jth activity is not performed.

In addition, allocation problems involving the allocation of men or machines give rise to IPP, since such commodities can be assigned only in integers and not in fractions.

Note: If the non-integer variable is rounded off, it violates the feasibility and there is no guarantee that the rounded off solution will be optimal. Due to these difficulties, there is a need for developing a systematic and efficient procedure for obtaining the exact optimal integer solution to such problems.

Types of Integer Programming Problem

There are two methods used to solve IPP, namely,

(*i*) Gomory's Cutting Plane Method

(*ii*) Branch and Bound Method (Search Method).

5.4.1 Gomory's All-IPP Method

A systematic procedure for solving pure IPP was first developed by R.E. Gomory, in 1956, which he later used to deal with the more complicated case of mixed integer programming problem. This method consists of first solving the IPP as an ordinary LPP by ignoring the restriction of integer values and then introducing a new constraint to the problem such that the new set of feasible solution includes all the original feasible integer solutions, but does not include the optimum non-integer solution initially found. This new constraint is called 'Fractional cut' or 'Gomorian constraint'. Then the revised problem is solved using the simplex method, till an optimum integer solution is obtained.

Search Method

This is an enumeration method in which all feasible integer points are enumerated. The widely used search method is the Branch and Bound method. It was developed in 1960, by A.H. Land and A.G. Doig. This method is applicable to both pure and mixed IPP. It first divides the feasible region into smaller subsets that eliminate parts containing no feasible integer solution.

5.4.2 Gomory's Fractional Cut Algorithm or Cutting Plane Method for Pure (All) IPP

Step 1 Convert the minimization IPP into an equivalent maximization IPP. Ignore the integrality condition.

Step 2 Introduce slack and/or surplus variables if necessary, to convert the given LPP in its standard form and obtain the optimum solution of the given LPP by using simplex method.

Step 3 Test the integrality of the optimum solution.

(*i*) If all $x_{Bi} \geq 0$ and are integers, an optimum integer solution is obtained.

(*ii*) If all $x_{Bi} \geq 0$ and at least one x_{Bi} is not an integer, then go to the next step.

Step 4 Rewrite each x_{Bi} as $x_{Bi} = [x_{Bi}] + f_i$ where x_{Bi} is the integral part of x_{Bi} and f_i is the positive fractional part of x_{Bi} $0 \leq f_i < 1$.

Choose the largest fraction of x_{Bi}'s, i.e., Choose max (f_i), if there is a tie, select arbitrarily. Let max $(f_i) = f_K$, corresponding to x_{BK} (the *K*th row is called the 'source row').

Step 5 Express each negative fraction, if any, in the source row of the optimum simplex table as the sum of a negative integer and a non-negative fraction.

Step 6 Find the fractional cut constraint (Gomorian Constraint)

From the source row $$\sum_{j=1}^{n} a_{k_j} x_j = x_{Bi}$$

i.e., $$\sum_{j=1}^{n} ([a_{kj}] + f_{kj})\, x_j = [x_{BK}] + f_K$$

in the form $$\sum_{j=1}^{n} f_{kj} x_j \geq f_K \quad -\sum_{j=1}^{n} f_{kj}\, x_j \leq -f_K$$

or, $$-\sum_{j=1}^{n} f_{kj} x_j + G_1 = -f_K$$

where, G_1 is the Gomorian slack.

Step 7 Add the fractional cut constraint obtained in step (6) at the bottom of the simplex table obtained in step (2). Find the new feasible optimum solution using dual simplex method.

Step 8 Go to step (3) and repeat the procedure until an optimum integer solution is obtained.

Example 5.19: Find the optimum integer solution to the following LPP.

Max $Z = x_1 + x_2$

Subject to constraints, $3x_1 + 2x_2 \leq 5$

$x_2 \leq 2$

$x_1, x_2 \geq 0$ and are integers.

Solution: Introducing the non-negative slack variable $S_1, S_2 \geq 0$, the standard form of the LPP becomes,

Max $Z = x_1 + x_2 + 0S_1 + 0S_2$

Subject to,

$$3x_1 + 2x_2 + S_1 = 5$$
$$0x_1 + x_2 + S_2 = 2$$
$$x_1, x_2, S_1, S_2 \geq 0$$

Ignoring the integrality condition, solve the problem by simplex method. The initial basic feasible solution is given by,

$$S_1 = 5 \text{ and } S_2 = 2.$$

Since all $Z_j - C_j \geq 0$ an optimum solution is obtained, given by

Max $Z = 7/3$, $x_1 = 1/3$, $x_2 = 2$.

To obtain an optimum integer solution, we have to add a fractional cut constraint in the optimum simplex table.

Since $x_B = 1/3$, the source row is the first row.

Expressing the negative fraction –2/3 as a sum of negative integer and positive fraction, we get

$$-2/3 = -1 + 1/3$$

	C_j		1	1	0	0	
C_B	B	x_B	x_1	x_2	S_1	S_2	$Min \frac{x_B}{x_1}$
← 0	S_1	5	(3)	2	1	0	5/3
0	S_2	2	0	1	0	1	—
	Z_j	0	0	0	0	0	
	$Z_j - C_j$		–1↑	–1	0	0	$Min \frac{x_B}{x_2}$
1	x_1	5/3	1	2/3	1/3	0	5/2 = 2 – 5
← 0	S_2	2	0	(1)	0	1	2/1 = 2
	Z_j	5/3	1	2/3	1/3	0	
	$Z_j - C_j$		0	–1/3↑	1/3	0	
1	x_1	1/3	1	0	1/3	–2/3	
1	x_2	2	0	1	0	1	
	Z_j	7/3	1	1	1/3	1/3	
	$Z_j - C_j$		0	0	1/3	1/3	

Since x_1 is the source row, we have,

$$1/3 = x_1 + 1/3\ S_1 - 2/3\ S_2$$

i.e.,
$$1/3 = x_1 + 1/3S_1 + (-1 + 1/3)\ S_2$$

The fractional cut (Gomorian) constraint is given by

$$1/3S_1 + 1/3S_2 \geq 1/3$$

$$\Rightarrow \quad -1/3S_1 - 1/3S_2 \leq -1/3$$

$$\Rightarrow \quad -1/3S_1 - 1/3\ S_2 + G_1 = -1/3$$

where, G_1 is the Gomorian slack. Add this fractional cut constraint at the bottom of the above optimal simplex table.

NOTES

		C_j	1	1	0	0	
C_B	B	x_B	x_1	x_2	S_1	S_2	G_1
1	x_1	1/3	1	0	1/3	–2/3	0
1	x_2	2	0	1	0	1	0
← 0	G_1	–1/3	0	0	(–1/3)	–1/3	1
	Z_j	7/3	1	1	1/3	1/3	0
	$Z_j - C_j$		0	0	1/3↑	1/3	0

We apply dual simplex method. Since $G_1 = -1/3$, G_1 leaves the basis. To find the entering variable we find,

$$\text{Max}\left\{\frac{Z_j - C_j}{a_{ik}}, a_{ik} < 0\right\} = \text{Max}\left\{\frac{1/3}{-1/3}, \frac{1/3}{-1/3}\right\}$$

$$\text{Max}\ \{-1, -1\} = -1$$

We choose S_1 as the entering variable arbitrarily.

		C_j	1	1	0	0	
C_B	B	x_B	x_1	x_2	S_1	S_2	G_1
1	x_1	0	1	0	0	–1	1
1	x_2	2	0	1	0	1	1
0	S_1	1	0	0	1	1	–3
	Z_j	2	1	1	0	0	1
	$Z_j - C_j$		0	0	0	0	1

Since all $Z_j - C_j \geq 0$ and all $x_{Bi} \geq 0$, we obtain an optimal feasible integer solution.

∴ The optimum integer solution is,

Max $Z = 2$, $x_1 = 0$, $x_2 = 2$.

Example 5.20: Find an optimum integer solution to the following LPP.

Max $Z = x_1 + 2x_2$

Subject to the contraints,

$$2x_2 \leq 7$$
$$x_1 + x_2 \leq 7$$
$$2x_1 \leq 11$$
$$x_1, x_2 \geq 0 \text{ and } x_1, x_2 \text{ are integers.}$$

Solution: Introducing slack variables $S_1, S_2, S_3 \geq 0$, we get,

Max $Z = x_1 + 2x_2 + 0S_1 + 0S_2 + 0S_3$

Subject to,

$$2x_2 + S_1 = 7$$
$$x_1 + x_2 + S_2 = 7$$
$$2x_1 + S_3 = 11$$

Ignoring the integer condition, we get the optimum solution of the given LPP, with initial basic feasible solution as, $S_1 = 7$, $S_2 = 7$, $S_3 = 11$.

	C_j		1	2	0	0	0	
C_B	B	x_B	x_1	x_2	S_1	S_2	S_3	$Min \frac{x_B}{x_2}$
←0	S_1	7	0	②	1	0	0	7/2 = 3.5
0	S_2	7	1	1	0	1	0	7/1 = 7
0	S_3	11	2	0	0	0	1	—
	Z_j	0	0	0	0	0	0	
	$Z_j - C_j$		–1	–2↑	0	0	0	$Min \frac{x_B}{x_1}$
2	x_2	7/2	0	1	1/2	0	0	—
←0	S_2	7/2	①	0	–1/2	1	0	7/2 = 3.5
0	S_3	11	2	0	0	0	1	11/2 = 5.5
	Z_j	7	0	2	1	0	0	
	$Z_j - C_j$		–1↑	0	1	0	0	
2	x_2	7/2	0	1	1/2	0	0	
1	x_1	7/2	1	0	–1/2	1	0	
0	S_3	4	0	0	1	–2	1	
	Z_j	21/2	1	2	1/2	1	0	
	$Z_j - C_j$		0	0	1/2	1	0	

Since all $Z_j - C_j \geq 0$, an optimum solution is obtained which is given by,

$$\text{Max Z} = \frac{21}{2}, x_1 = \frac{7}{2}, x_2 = \frac{7}{2}$$

Since the optimum solution obtained above is not an integer, we now select a constraint corresponding to

$$\text{Max } \{fi\} = \text{Max } \{f_1, f_2, f_3\}$$

$$x_1 = 7/2 = 3 + 1/2$$

$$x_2 = 7/2 = 3 + 1/2$$

$$S_3 = 4 = 4 + 0$$

$\therefore$ $$\text{Max } \{f_i\} = \text{Max } \left(\frac{1}{2}, \frac{1}{2}, 0\right) = 1/2$$

Since the max fraction is same for both x_1 and x_2 rows, we choose x_1 row as the source row arbitrarily. From this row we have,

$$7/2 = x_1 + 0x_2 - 1/2\ S_1 + 1S_2 + 0S_3.$$

On expressing the negative fraction as a sum of negative integer and a positive fraction, we have,

$$3 + 1/2 = x_1 + 0x_2 + (-1 + 1/2)\ S_1 + 1S_2 + 0S_3$$

$\therefore$ The Gomorian constraint is given by,

$$1/2\ S_1 \geq 1/2$$

i.e., $$-1/2\ S_1 \leq -1/2 \Rightarrow -1/2\ S_1 + G_1 = -1/2$$

where, G_1 is the Gomorian slack. Adding this new constraint at the bottom of the above optimal simplex table, we get a new table.

NOTES

	C_j		1	2	0	0	0	0
C_B	B	x_B	x_1	x_2	S_1	S_2	S_3	G_1
2	x_2	7/2	0	1	1/2	0	0	0
1	x_1	7/2	1	0	−1/2	1	0	0
0	S_3	4	0	0	1	−2	1	0
←0	G_1	−1/2	0	0	(−1/2)	0	0	1
	Z_j	21/2	1	2	1/2	1	0	0
	$Z_j - C_j$		0	0	1/2	1	0	0

We apply dual simplex method. Since $G_1 = -1/2$, G_1 leaves the basis. Entering variable is given by,

$$\text{Max}\left\{\frac{Z_j - C_j}{a_{k_j}}, a_{k_j} < 0\right\} = \text{Max}\left\{\frac{1/2}{-1/2}\right\}$$

gives the non-basic variable S_1 to enter into the basis. Drop G_1 and introduce S_1.

	C_j		1	2	0	0	0	0
C_B	B	x_B	x_1	x_2	S_1	S_2	S_3	G_1
2	x_2	3	0	1	0	0	0	1
1	x_1	4	1	0	0	1	0	−1
0	S_3	3	0	0	0	−2	1	2
0	S_1	1	0	0	1	0	0	−2
	Z_j	10	1	2	0	1	0	1
	$Z_j - C_j$		0	0	0	1	0	1

Since all $Z_j - C_j \geq 0$, an optimum solution has been obtained in integers. Hence, the integer optimum solution is given by,

$$\text{Max } Z = 10,\ x_1 = 4,\ x_2 = 3.$$

Example 5.21: Solve the following integer programming problem.

Max $Z = 2x_1 + 20x_2 - 10x_3$

Subject to, $2x_1 + 20x_2 + 4x_3 \leq 15$

$6x_1 + 20x_2 + 4x_3 = 20$

$x_1, x_2, x_3 \geq 0$ and are integers.

Solution: Introducing slack variable $S_1 \geq 0$ and an artificial variable $A_1 \geq 0$, the initial basic feasible solution is $S_1 = 15, A_1 = 20$. Ignoring the integer condition, solve the problem by simplex method.

Max $Z = 2x_1 + 20x_2 - 10x_3 + 0S_1 - MA_1$

Subject to, $2x_1 + 20x_2 + 4x_3 + S_1 = 15$

$6x_1 + 20x_2 + 4x_3 + A_1 = 20$

$x_1, x_2, x_3, S_1, A_1 \geq 0$

	C_j		2	20	−10	0	−M	
C_B	B	x_B	x_1	x_2	x_3	S_1	A_1	Min $\frac{x_B}{x_2}$
←0	S_1	15	2	(20)	4	1	0	15/20 = 3/4
−M	A_1	20	6	20	4	0	1	20/20 = 1
	Z_j	−20M	−6M	−20M	−4M	0	−M	
	$Z_j - C_j$		−6M − 2	−20M − 20 ↑	−4M + 10	0	0	

	C_j		2	20	−10	0	−M	
C_B	B	x_B	x_1	x_2	x_3	S_1	A_1	Min x_B/x_1
20	x_2	3/4	1/10	1	$\frac{1}{5}$	1/20	0	$\frac{3}{4}\times 10 = \frac{15}{2} = 7.5$
← −M	A_1	5	4	0	0	−1	1	$\frac{5}{4} = 1.25$
	Z_j	15 − 5M	2 − 4M	20	4	1 + M	−M	
	$Z_j - C_j$		−4 — ↑	0	14	M + 1	0	
20	x_2	5/8	0	1	1/5	3/40	—	
2	x_1	5/4	1	0	0	−1/4	—	
	Z_j	15	2	20	4	1	—	
	$Z_j - C_j$		0	0	14	1		

Since all $Z_j - C_j \geq 0$, the solution is optimum but the variables are non-integer.

∴ The non-integer optimum solution is given by,

$$x_1 = 5/4, x_2 = 5/8, x_3 = 0, \text{ Max } Z = 15$$

To obtain an integer optimum solution, we proceed as follows.

$$\text{Max } \{f_1 f_2\} = \text{Max } \{5/8, 1/4\} = 5/8$$

∴ The source row is the first row, namely, x_2 row. From this source row we have,

$$5/8 = 0x_1 + 1x_2 + (1/5)x_3 + (3/40)S_1.$$

The fractional cut constraint is given by,

$$(1/5)x_3 + (3/40)S_1 \geq 5/8$$

$$(-1/5)x_3 - (3/40)S_1 \leq -5/8 \Rightarrow (-1/5)x_3 - (3/40)S_1 + G_1 = 5/8$$

where, G_1 is the Gomorian slack.

	C_j		2	20	−10	0	0
C_B	B	x_B	x_1	x_2	x_3	S_1	G_1
20	x_2	$\frac{5}{8}$	0	1	1/5	3/40	0
2	x_1	$\frac{5}{4}$	1	0	0	−1/4	0
←0	G_1	$\frac{-5}{8}$	0	0	−1/5	(−3/40)	1
	Z_j	15	2	20	4	1	0
	$Z_j - C_j$		0	0	14	1 ↑	0

NOTES

We apply dual simplex method. Since $G_1 = -5/8$, G_1 leaves the basis.

Also, $$\text{Max}\left\{\frac{Z_j - C_j}{a_i k}, a_i k < 0\right\} = \text{Max}\left\{\frac{14}{-1/5}\ \frac{1}{-3/40}\right\} = \text{Max} -\frac{40}{3}$$

gives the non-basic variable S_1, this enters the basis.

	c_j		2	20	–10	0	0
C_B	B	x_B	x_1	x_2	x_3	S_1	G_1
20	x_2	0	0	1	0	0	+1
2	x_1	10/3	1	0	2/3	0	–10/3
0	S_1	25/3	0	0	8/3	1	–40/3
	z_j	20/3	2	20	4/3	0	40/3
	$z_j - c_j$		0	0	34/3	0	40/3

Again since the solution is non-integer, we add one more fractional cut constraint.

$$\text{Max}\{f_i\} = \text{Max}\ \{0, 1/3, 1/3\}$$

Since the max fraction is same for both the rows x_1 and S_1, we choose S_1 arbitrarily.

∴ From the source row we have,

$$25/3 = 0x_1 + 0x_2 + (8/3)x_3 + 1S_1 - (40/3)G_1$$

Expressing the negative fraction as the sum of negative integer and positive fraction we have,

$$(8 + 1/3) = 0x_1 + 0x_2 + (2 + 2/3)x_3 + 1S_1 + (-14 + 2/3)G_1$$

The corresponding fractional cut is given by,

$$-2/3x_3 - 2/3\ G_1 + G_2 = -1/3.$$

Add this second Gomorian constraint at the bottom of the above simplex table and apply dual simplex method.

	c_j		2	20	–10	0	0	0
C_B	B	x_B	x_1	x_2	x_3	S_1	G_1	G_2
20	x_2	0	0	1	0	0	1	0
2	x_1	10/3	1	0	2/3	0	–10/3	0
0	S_1	25/3	0	0	8/3	1	–40/3	0
← 0	G_2	–1/3	0	0	(–2/3)	0	–2/3	1
	z_j	20/3	2	20	4/3	0	40/3	0
	$z_j - c_j$		0	0	34/3↑	0	40/3	0

Since $G_2 = -1/3$, G_2 leaves the basis. Also,

$$\text{Max}\left(\frac{Z_j - C_j}{a_i k}, a_i k < 0\right) = \text{Max}\left(\frac{34/3}{-2/3} - \frac{40/3}{-2/3} -\right) = -17$$

gives the non-basic variable x_3 which enters the basis. Using dual simplex method, introduce x_3 and drop G_2.

NOTES

	C_j		2	20	–10	0	0	0
C_B	B	x_B	x_1	x_2	x_3	S_1	G_1	G_2
20	x_2	0	0	1	0	0	1	0
2	x_1	3	1	0	0	0	–4	1
0	S_1	7	0	0	0	1	16	4
–10	x_3	1/2	0	0	1	0	1	–3/2
	Z_j	1	2	20	–10	0	2	17
	$Z_j - C_j$		0	0	0	0	2	17

Since the solution is still a non-integer, a third fractional cut is required. It is given from the source row (x_3 row) as,

$$-1/2 = -1/2\ G_2 + G_3$$

Insert this additional constraint at the bottom of the table, the modified simplex table is shown below.

	C_j		2	20	–10	0	0	0	0
C_B	B	x_B	x_1	x_2	x_3	S_1	G_1	G_2	G_3
20	x_2	0	0	1	0	0	1	0	0
2	x_1	3	1	0	0	0	–4	1	0
0	S_1	7	0	0	0	1	–16	4	0
–10	x_3	1/2	0	0	1	0	1	3/2	0
←0	G_3	–1/2	0	0	0	0	0	–1/2	1
	Z_j	1	2	20	–10	0	2	17	0
	$Z_j - C_j$		0	0	0	0	2	17↑	0

Using dual simplex method, we drop G_3 and introduce G_2.

	C_j		2	20	–10	0	0	0	0
G_B	B	x_B	x_1	x_2	x_3	S_1	G_1	G_2	G_3
20	x_2	0	0	1	0	0	0	0	0
2	x_1	2	1	0	0	0	–4	0	2
0	S_1	3	0	0	0	1	–16	0	8
–10	x_3	2	0	0	1	0	–1	0	–3
0	G_2	1	0	0	0	0	6	1	–2
	Z_j	–16	2	20	–10	0	2	0	34
	$Z_j - C_j$		0	0	0	0	2	0	34

Since all $Z_j - C_j \geq 0$ and also the variables are integers, the optimum integer solution is obtained and given by, $x_1 = 2, x_2 = 0, x_3 = 2$ and Max $Z = 16$.

Example 5.22: Solve the integer programming problem.

Max $Z = 7x_1 + 9x_2$

Subject to, $-x_1 + 3x_2 \leq 6$

$7x_1 + x_2 \leq 35$

$x_1, x_2 \geq 0$ are integers.

Solution: Introducing slack variables $S_1, S_2 \geq 0$, we get the standard form of LPP as,

Max $Z = 7x_1 + 9x_2 + 0S_1 + 0S_2$

Subject to, $-x_1 + 3x_2 + S_1 = 6$

$$7x_1 + x_2 + S_2 = 35$$

Now ignoring the integer conditions, solve the given LPP by simplex method.

		C_j	7	9	0	0	
C_B	B	x_B	x_1	x_2	S_1	S_2	$Min \frac{x_B}{x_2}$
←0	S_1	6	–1	(3)	1	0	6/3 = 2
0	S_2	35	7	1	0	1	35/1 = 35
	Z_j	0	0	0	0	0	
	$Z_j - C_j$		–7	–9↑	0	0	
9	x_2	2	–1/3	1	1/3	0	–
←0	S_2	33	(22/3)	0	–1/3	1	$33\times\frac{22}{2}$
	Z_j	18	–3	9	3	0	
	$Z_j - C_j$		–10↑	0	3	0	
9	x_2	7/2	0	1	7/22	1/22	
7	x_1	9/2	1	0	–1/22	3/22	
	Z_j	63	7	9	28/11	15/11	
	$Z_j - C_j$		0	0	28/11	15/11	

Since all $Z_j - C_j \geq 0$, optimum solution is obtained as $x_1 = \frac{9}{2}$

$$x_2 = \frac{7}{2} \text{ and Max } Z = 63.$$

Since the optimum solution obtained above is not an integer solution, we select a constraint corresponding to,

$$\text{Max } \{f_i\} = \text{Max } \{f_1 f_2\}$$

$$= \text{Max} \left\{\frac{1}{2}, \frac{1}{2}\right\} \left[x_{B1} = \frac{7}{2} = [3] + \frac{1}{2}, x_{B2} = \frac{9}{2} = [4] + \frac{1}{2}\right]$$

Since both the equations have the same value of f_i, either one of the two equations can be used. Let us consider the x_2 row as source row.

From x_2 row we have,

$$\frac{7}{2} = 0x_1 + x_2 + \frac{7}{22}S_1 + \frac{1}{22}S_2$$

There is no negative fraction.

The Gomorian constraint is given by,

$$\frac{7}{22}S_1 + \frac{1}{22}S_2 \geq \frac{1}{2}$$

i.e., $$\frac{7}{22}S_1 - \frac{1}{22}S_2 \leq -\frac{1}{2}$$

$\Rightarrow$ $$-\frac{7}{22}S_1 - \frac{1}{22}S_2 + G_1 = -\frac{1}{2}$$

where, G_1 is the Gomorian slack. Adding this new constraint at the bottom of the above optimal simplex table, we have the new table.

	C_j		7	9	0	0	0
C_B	B	x_B	x_1	x_2	S_1	S_2	G_1
9	x_2	9/2	0	1	7/22	1/22	0
7	x_1	7/2	1	0	–1/22	3/22	0
0	G_1	–1/2	0	0	–7/22	–1/22	1
	Z_j	63	9	7	28/11	15/11	0
	$Z_j - C_j$		0	0	28/11↑	15/11	0

We apply dual simplex method, since $G_1 = -1/2$, G_1 leaves the basis. Also,

$$\text{Max}\left\{\frac{Z_j - C_j}{a_{ik}}, a_{ik} < 0\right\} = \text{Max}\left\{\frac{\frac{28}{11}}{-\frac{7}{22}}, \frac{\frac{15}{11}}{-\frac{1}{22}}\right\}$$

$$= \text{MAX}\ (-8, -30) = -8$$

gives the non-basic variable S_1 to enter into the basis.

Applying dual simplex method, drop G_1 and introduce S_1.

	C_j		7	9	0	0	0
C_B	B	x_B	x_1	x_2	S_1	S_2	G_1
9	x_2	3	0	1	0	0	1
7	x_1	32/7	1	0	0	1/7	–1/7
0	S_1	11/7	0	0	1	+1/7	–22/7
	Z_j	59	7	9	0	1	8
	$Z_j - C_j$		0	0	0	1	8

The optimal solution obtained by dual simplex method as above is still a non-integer. Thus a new Gomory's constraint is to be reconsidered.

$$\text{Max } \{f_i\} = \text{Max}\left\{-, \frac{4}{7}, \frac{4}{7}\right\} = \frac{4}{7}$$

Choose the x_1 row as source row arbitrarily as both the fraction values are the same. From the source row we have,

$$\frac{4}{7} = 1x_1 + 0x_2 + 0S_1 + \frac{1}{7}S_2 + \frac{6}{7}G_1$$

There is no negative fraction in the source row.

The Gomory's constraint is given by,

$$\frac{1}{7}S_2 + \frac{6}{7}G_1 \geq \frac{4}{7} \quad \text{i.e.,} \quad \frac{1}{7}S_2 - \frac{6}{7}G_1 + G_2 = -\frac{4}{7}$$

where, G_2 is the Gomorian slack. Adding this constraint in the above simplex table we get a modified table.

	C_j		7	9	0	0	0	0
C_B	B	x_B	x_1	x_2	S_1	S_2	G_1	G_2
9	x_2	3	0	1	0	0	1	0
7	x_1	32/7	1	0	0	1/7	–1/7	0
0	S_1	11/7	0	0	1	1/7	–22/7	0
0	G_2	–4/7	0	0	0	–1/7	–6/7	1
	Z_j	59	7	9	0	1	8	0
	$Z_j - C_j$		0	0	0	1↑	8	0

We again apply the dual simplex method.

Since $G_2 = -\frac{4}{7}$, G_2 leaves the basis. Also,

$$\text{Max}\left\{\frac{Z_j - C_j}{a_{ik}}, a_{ik} < 0\right\} = \text{Max}\left\{\frac{1}{-\frac{1}{7}}, \frac{8}{-\frac{6}{7}}\right\}$$

$$= \text{Max}\,(-7, -9) = -7$$

gives the non-basic variable S_2 to enter into the basis.

	C_j		7	9	0	0	0	0
C_B	B	x_B	x_1	x_2	S_1	S_2	G_1	G_2
9	x_2	3	0	1	0	0	1	0
7	x_1	4	1	0	0	0	–1	1
0	S_1	1	0	0	1	0	–4	1
0	S_2	4	0	0	0	1	6	–7
	Z_j	55	7	9	0	0	2	7
	$Z_j - C_j$		0	0	0	0	2	7

Since all $Z_j - C_j \geq 0$ and also the solution is an integer, we obtain an optimum integer solution given by, $x_1 = 4, x_2 = 3$ and Max $Z = 55$.

Mixed Integer Programming Problem

In mixed IPP only some of the variables are restricted to integer values, while the other variables may take integer or other real values.

Mixed integer cutting plane procedure The iterative procedure for the solution of mixed integer programming problem is as follows.

Step 1 Reformulate the given LPP into a standard maximization form and then determine an optimum solution using simplex method.

Step 2 Test the integrality of the optimum solution.

(*i*) If all $x_{Bi} \geq 0$ ($i = 1, 2,\ldots m$) and are integers, then the current solution is an optimum one.

(*ii*) If all $x_{Bi} \geq 0$ ($i = 1, 2,\ldots m$) but the integer restricted variables are not integers, then go to the next step.

Step 3 Choose the largest fraction among those x_{Bi}, which are restricted to integers. Let it be $x_{Bk} = f_k$ (assume)

Step 4 Find the fractional cut constraints from the source row, namely *K*th row.

From the source row,

$$\sum_{j=1}^{n} a_{kj}\, k_j = x_{Bk}$$

i.e,

$$\sum_{j=1}^{n}\left(\left[a_{kj}\right] + f_{ki}\right) r_j = [x_{BK}] + f_k$$

in the form

$$\sum_{j=1}^{n} f_{ki}\, x_j \geq f_k$$

i.e., $$\sum_{j \in j^+} f_{kj} x_j + \left(\frac{f_k}{f_{k-1}}\right) \sum_{j \in j^-} f_{kj} x_j \geq + f_k$$

$$-\sum_{j \in j^+} f_{kj} x_j - \left(\frac{f_k}{f_{k-1}}\right) \sum_{j \in j^-} f_{kj} x_j \leq - f_k$$

$$-\sum_{j \in j^+} f_{kj} x_j - \left(\frac{f_k}{f_{k-1}}\right) \sum_{j \in j^-} f_{kj} x_j + G_k = - f_k$$

where, G_k is Gomorian slack

$$\begin{matrix} j^+ = \\ j^{-1} = \end{matrix} \begin{bmatrix} j / f_{kj} \geq 0 \\ j / f_{kj} < 0 \end{bmatrix}$$

Step 5 Add this cutting plane generated in step K at the bottom of the optimum simplex table obtained in step 1. Find the new optimum solution using dual simplex method.

Step 6 Go to step 2 and repeat the procedure until all $x_{Bi} \geq 0$ $(i = 1, 2, \ldots m)$ and all restricted variables are integers.

Example 8.23:

Max $Z = x_1 + x_2$

Subject to, $3x_1 + 2x_2 \leq 5$

$x_2 \leq 2$

$x_1 + x_2 \geq 0$ and x_1 is an integer.

Solution: Introducing slack variables $S_1, S_2 \geq 0$ the standard form of LPP is,

Max $Z = x_1 + x_2 + 0S_1 + 0S_2$

Subject to, $3x_1 + 2x_2 + S_1 = 5$

$x_2 + S_2 = 2$

$x_1, x_2, S_1, S_2 \geq 0$

Initial basic feasible solution,

$$S_1 = 5, S_2 = 2$$

Ignore the integer condition and solve the problem using simplex method, to obtain optimum solution.

	C_j		1	1	0	0	
C_B	B	x_B	x_1	x_2	S_1	S_2	$Min \frac{x_B}{x_1}$
←0	S_1	5	(3)	2	1	0	5/3
0	S_2	2	0	1	0	1	—
	Z_j	0	0	0	0	0	
	$Z_j - C_j$		−1↑	−1	0	0	$Min \frac{x_B}{x_2}$
1	x_1	5/3	1	2/3	1/3	0	5/2
←0	S_2	2	0	(1)	0	1	5/2
	Z_j	5/3	1	2/3	1/3	0	
	$Z_j - C_j$		0	−1/3↑	1/3	0	
1	x_1	1/3	1	0	1/3	−2/3	
1	x_2	2	0	1	0	1	
	Z_j	7/3	1	1	1/3	1/3	
	$Z_j - C_j$		0	0	1/3	1/3	

Since all $Z_j - C_j \geq 0$, the current basic feasible solution is optimum. But x_1 is non-integer. From the source row (first row) we have,

$$1/3 = x_1 + 0\,x_2 + 1/3\,S_1 - 2/3\,S_2$$

The Gomorian constraint is given by,

$$\frac{1}{3}S_1 + \left(\frac{\frac{1}{3}}{\frac{1}{3}-1}\right)\left(\frac{-2}{3}\right)S_2 \geq \frac{1}{3}$$

$$\frac{1}{3}S_1 + \frac{1}{3}S_2 \geq \frac{1}{3} \quad \Rightarrow \quad -\frac{1}{3}S_1 - \frac{1}{3}S_2 \leq -\frac{1}{3}$$

$$\frac{-1}{3}S_1 - \frac{1}{3}S_2 + G_1 = -\frac{1}{3}$$

where, G_1 is the Gomorian slack.

Adding this Gomorian constraint at the bottom of the above simplex table, we have,

	C_j		1	1	0	0	0
C_B	B	x_B	x_1	x_2	S_1	S_2	G_1
1	x_1	1/3	1	0	1/3	–2/3	0
1	x_2	2	0	1	0	1	0
0	G_1	–1/3	0	0	(–1/3)	–1/3	1
	Z_j	7/3	1	1	1/3	1/3	0
	$Z_j - C_j$		0	0	1/3↑	1/3	0

Using the dual simplex method, since $G_1 = -1/3 < 0$, G_1 leaves the basis. Also,

$$\text{Max}\left\{\frac{Z_j - C_j}{a_{ik}},\ a_{ik} < 0\right\}$$

$$\text{Max}\left\{\frac{\frac{1}{3}}{\frac{-1}{3}}, \frac{\frac{1}{3}}{\frac{-1}{3}}\right\} = \text{Max}\ \{-1, -1\} = -1$$

As this corresponds to both S_1 and S_2, we choose S_1 arbitrarily as the entering variable.

Drop G_1 and introduce S_1.

	C_j		1	1	0	0	0
C_B	B	x_B	x_1	x_2	S_1	S_2	G_1
1	x_1	0	1	0	0	- 1	1
1	x_2	2	0	1	0	1	0
0	S_2	1	0	0	1	1	- 3
	Z_j	2	1	1	0	0	1
	$Z_j - C_j$		0	0	0	0	1

Since all $Z_j - C_j \geq 0$ and all $x_{Bi} \geq 0$, the current solution is feasible and optimal.

The required optimal integer solution is given by,

$$x_1 = 0,\ x_2 = 2 \text{ and Max } Z = 2.$$

Example 5.24:

Max $Z = 4x_1 + 6x_2 = 2x_3$

Subject to,

$$4x_1 - 4x_2 \leq 5,$$
$$-x + 6x_2 \leq 5,$$
$$-x_1 + x_2 + x_3 \leq 5$$
$$x_1, x_2, x_3 \geq 0 \text{ and } x_1, x_3 \text{ are integers.}$$

Solution Introducing slack variables $S_1, S_2, S_3 \geq 0$ the standard form of LPP is,

Max $Z = 4x_1 + 6x_2 + 2x_3 + 0S_1 + 0S_2 + 0S_3$

Subject to,

$$4x_1 - 4x_2 + S_1 = 5$$
$$-x_1 + 6x_2 + S_2 = 5$$
$$-x_1 + x_2 + x_3 + S_3 = 5$$

The initial basic feasible solution is given by $S_1 = 5, S_2 = 5$ and $S_3 = 5$. Ignoring the integer condition, the optimum solution of given LPP is obtained by the simplex method.

Since all $Z_j - C_j \geq 0$, the solution is optimum. But the integer constrained variables x_1 and x_3 are non-integer.

$$x_1 = 5/2 = 2 + 1/2$$
$$x_2 = 25/4 = 6 + 1/4$$

Max $(f_1\, f_3)$ = Max (1/2, 1/4) = 1/2

From the first row we have,

$$(2 + 1/2) = x_1 + 0x_2 + 0x_3 + (3/10)\, S_1 + (1/5)\, S_2$$

	c_j		4	6	2	0	0	0	
C_B	B	x_B	x_1	x_2	x_3	S_1	S_2	S_3	Min x_B/x_2
0	S_1	5	4	–4	0	1	0	0	—
←0	S_2	5	–1	(6)	0	0	1	0	5/6
0	S_3	5	–1	1	1	0	0	1	5/1
	Z_j	0	0	0	0	0	0	0	
	$Z_j - C_j$		–4	–6↑	–2	0	0	0	Min x_B/x_1
0	S_1	25/3	(10/3)	0	0	1	2/3	0	25/10
6	x_2	5/6	–1/6	1	0	0	1/6	0	—
0	S_3	25/6	–5/6	0	1	0	–1/6	1	—
	Z_j	5	–1	6	0	0	1	0	
	$Z_j - C_j$		–5↑	0	–2	0	1	0	Min x_B/x_3
4	x_1	5/2	1	0	0	3/10	1/5	0	—
6	x_2	5/4	0	1	0	1/20	1/5	0	25
←0	S_3	25/4	0	0	(1)	1/4	0	1	4/1
	Z_j	35/2	4	6	0	3/2	2	0	
	$Z_j - C_j$		0	0	–2↑	3/2	2	0	
4	x_1	5/2	1	0	0	3/10	1/5	0	
6	x_2	5/4	0	1	0	1/20	1/5	0	
2	x_3	25/4	0	0	1	1/4	0	1	
	Z_j	35/2	4	6	2	3/2	2	0	
	$Z_j - C_j$		0	0	0	3/2	2	0	

NOTES

The Gomorian constraint is given by,

$$3/10\ S_1 + 1/5\ S_2 \geq 1/2$$

$$-3/10\ S_1 - 1/5\ S_2 \leq -1/2$$

i.e., $-3/10\ S_1 - 1/5\ S_2 + G_1 = -1/2$, where G_1 is the Gomorian slack. Introduce this new constraint at the bottom of the above simplex table.

Using dual simplex method, since $G_1 = -1/2 < 0$,

G_1 leaves the basis. Also,

$$\text{Max}\left\{\frac{Z_j - C_j}{a_{ik}}, a_{ik} < 0\right\} = \text{Max}\left\{\frac{2}{\frac{-3}{10}}, \frac{2}{\frac{-1}{5}}\right\} = \text{Max}\left\{\frac{-20}{3}, -10\right\} = \frac{-20}{3}$$

		C_j	4	6	2	0	0	0
C_B	B	x_B	x_1	x_2	x_3	S_1	S_2	G_1
4	x_1	5/2	1	0	0	3/10	1/5	0
6	x_2	5/4	0	1	0	1/20	1/5	0
2	x_3	25/4	0	0	1	1/4	0	0
←0	G_1	−1/2	0	0	0	(−3/10)	−1/5	1
	Z_j	30	4	6	2	2	2	0
	$Z_j - C_j$		0	0	0	2↑	2	0

corresponding to S_1. Therefore, the non-basic variable S_1 enters the basics. Drop G_1 and introduce S_1.

		C_j	4	6	2	0	0	0	
C_B	B	x_B	x_1	x_2	x_3	S_1	S_2	S_3	G_1
4	x_1	2	1	0	0	0	0	0	1
6	x_2	7/6	0	1	0	0	1/6	0	1/6
2	x_3	35/6	0	0	1	0	−1/6	1	5/6
0	S_1	5/3	0	0	0	1	2/3	0	−10/3
	Z_j	80/3	4	6	2	0	2/3	2	20/3
	$Z_j - C_j$		0	0	0	0	2/3	2	20/3

Since all $Z_j - C_j \geq 0$, the solution is optimum and also the integer restricted variable $x_3 = 35/6$ is not an integer, therefore, we add another Gomorian constraint

$$x_3 = 35/6 = 5 + 5/6$$

The source row is the third row.

From this row we have,

$$5 + \frac{5}{6} = 0x_1 + 0x_2 + x_3 + 0S_1 - \frac{1}{6}S_2 + S_3 + \frac{5}{6}G_1$$

The Gomorian constraint is given by,

$$\left(\frac{\frac{5}{6}}{\frac{5}{6}-1}\right)\left(\frac{-1}{6}\right)S_2 + \frac{5}{6}G_1 \geq \frac{5}{6}$$

$$\Rightarrow \qquad \frac{5}{6}S_2 + \frac{5}{6}G_1 \geq \frac{5}{6}$$

$$\Rightarrow \qquad \frac{-5}{6}S_2 - \frac{5}{6}G_1 + G_2 \leq \frac{-5}{6}$$

where, G_2 is the Gomorian slack.

Add this second cutting plane constraint at the bottom of the above optimum simplex table.

	c_j		4	6	2	0	0	0	0	0
C_B	B	x_B	x_1	x_2	x_3	S_3	S_1	S_2	G_1	G_2
4	x_1	2	1	0	0	0	0	0	1	0
6	x_2	7/6	0	1	0	0	0	1/6	1/6	0
2	x_3	35/6	0	0	1	1	0	–1/6	5/6	0
0	S_1	5/3	0	0	0	0	1	2/3	–10/3	0
← 0	G_2	–5/6	0	0	0	0	0	–5/6	–5/6	1
	Z_j	80/3	4	6	2	2	2/3	2/3	20/3	0
	$Z_j - C_j$		0	0	0	2	2/3	↑2/3	20/3	0

Use dual simplex method. ($\because G_2 = -5/6 < 0$)

G_2 leaves the basics.

$$\text{Also, Max}\left\{\frac{Z_j - C_j}{a_{ik}}\; a_{ik} < 0\right\} = \text{Max}\left\{\frac{\frac{2}{3}}{\frac{-5}{6}}, \frac{\frac{20}{3}}{\frac{-5}{6}}\right\} = \text{Max}\left\{\frac{-4}{5}, -8\right\} = -\frac{4}{5},$$

which corresponds to S_2.

Drop G_2 and introduce S_2.

Since all $Z_j - C_j \geq 0$ and also all the restricted variables x_1 and x_3 are integers, an optimum integer solution is obtained.

The optimum integer solution is,

$x_1 = 2$, $x_2 = 1$, $x_3 = 6$ and Max $Z = 26$

	c_j		4	6	2	0	0	0	0	0
C_B	B	x_B	x_1	x_2	x_3	S_1	S_2	S_3	G_1	G_2
4	x_1	2	1	0	0	0	0	0	1	0
6	x_2	1	0	1	0	0	0	0	0	1/5
2	x_3	6	0	0	1	0	0	1	1	–1/5
0	S_1	1	0	0	0	1	0	0	–4	4/5
0	S_2	1	0	0	0	0	1	0	1	–6/5
	Z_j	26	4	6	2	0	0	2	6	4/5
	$Z_j - C_j$		0	0	0	0	0	2	6	4/5

5.4.3 Branch and Bound Technique

This method is applicable to both, pure as well as mixed IPP. Sometimes a few or all the variables of an IPP are constrained by their upper or lower bounds.

The most general method for the solution of such constrained optimization problems is called 'Branch and Bound method'.

This method first divides the feasible region into smaller subsets and then examines each of them successively, until a feasible solution that gives an optimal value of objective function is obtained.

Let the given IPP be,

Max $Z = CX$

Subject to, $Ax \leq b$

$X \geq 0$ are integers.

In this method, we first solve the problem by ignoring the integrality condition.

(*i*) If the solution is in integers, the current solution is optimum for the given IPP.

(*ii*) If the solution is not in integers, say one of the variable X_r is not an integer, then $x_r^* < x_r < x^*_{r+1}$ where x^*_r, x^*_{r+1} are consecutive non-negative integers.

Hence, any feasible integer value of x_r must satisfy one of the two conditions.

$$x_r \leq x_r^* \text{ or } x_r \geq x^*_{r+1}.$$

These two conditions are mutually exclusive (both cannot be true simultaneously). By adding these two conditions separately to the given LPP, we form different sub-problems.

Sub-problem 1	Sub-problem 2
Max $Z = Cx$	Max $Z = Cx$
Subject to, $Ax \leq b$	Subject to, $Ax \leq b$
$x_r \leq x^*_r$	$x_r \geq x^*_{r+1}$
$x \geq 0.$	$x \geq 9.$

Thus, we have branched or partitioned the original problem into two sub-problems. Each of these sub-problems is then solved separately as LPP.

If any sub-problem yields an optimum integer solution, it is not further branched. But if any sub-problem yields a non-integer solution, it is further branched into two sub-problems. This branching process is continued until each problem terminates with either an integer optimal solution or there is an evidence that it cannot yield a better solution. The integer-valued solution among all the sub-problems, which gives the most optimal value of the objective function is then selected as the optimum solution.

Note: For minimization problem, the procedure is the same except that upper bounds are used. The sub-problem is said to be fathomed and is dropped from further consideration if it yields a value of the objective function lower than that of the best available integer solution and it is useless to explore the problem any further.

Example 5.25: Use branch and bound technique to solve the following:

Max $Z = x_1 + 4x_2$

Subject to, $2x_1 + 4x_2 \leq 7$

$$5x_1 + 3x_2 \leq 15$$

$$x_1, x_2 \geq 0 \text{ and are integers.}$$

Solution: Ignoring the integrality condition we solve the LPP,

NOTES

Max $Z = x_1 + 4x_2$

Subject to, $2x_1 + 4x_2 \leq 7$

$5x_1 + 3x_2 \leq 15$

$x_1, x_2 \geq 0$

Introducing slack variables $S_1, S_2 \geq 0$, the standard form of LPP becomes,

Max $Z = x_1 + 4x_2 + 0S_1 + 0S_2$

Subject to, $2x_1 + 4x_2 + S_1 = 7$

$5x_1 + 3x_2 + S_2 = 15$

		C_j	1	4	0	0	
C_B	B	x_B	x_1	x_2	S_1	S_2	Min x_B/x_2
←0	S_1	7	2	(4)	1	0	7/4
0	S_2	15	5	3	0	1	15/3 = 5
	Z_j	0	0	0	0	0	
	$Z_j - C_j$		–1	–4↑	0	0	
4	x_2	7/4	$\frac{1}{2}$	1	1/4	0	
0	S_2	$\frac{39}{4}$	$\frac{7}{2}$	0	–3/4	1	
	Z_j	7	2	4	1	0	
	$Z_j - C_j$		1	0	1	0	

Since all $Z_j - C_j \geq 0$, an optimum solution is obtained.

$x_1 = 0$, $x_2 = 7/4$ and Max $Z = 7$

Since $x_2 = \frac{7}{4}$, this problem should be branched into two sub-problems.

For $x_2 = \frac{7}{4}$, $1 < x_2 < 2 = x_2 \leq 1,\ x_2 \geq 2$

Applying these two conditions separately in the given LPP we get two sub-problems.

Sub-problem (1)

Max $Z = x_1 + 4x_2$

Subject to, $2x + 4x_2 \leq 7$

$5x_1 + 3x_2 \leq 15$

$x_2 \leq 1$

$x_1, x_2 \geq 0$

Sub-problem (2)

Max $Z = x_1 + x_2$

Subject to, $2x_1 + 4x_2 \leq 7$

$5x_1 + 3x_2 \leq 15$

$x_2 \geq 2$

$x_1, x_2 \geq 0$

NOTES

Sub-Problem (1)

C_B	B	x_B	x_1	x_2	S_1	S_2	S_3	Min $\frac{x_B}{x_2}$
	C_j		1	4	0	0	0	
0	S_1	7	2	4	1	0	0	7/4
0	S_2	15	5	3	0	1	0	15/3
←0	S_3	1	0	(1)	0	0	1	1/1
	Z_j	0	0	0	0	0	0	
	$Z_j - C_j$		–1	–4↑	0	0	0	Min $\frac{x_B}{x_1}$
←0	S_1	3	(2)	0	1	0	–4↑	3/2
0	S_2	12	5	0	0	1	–3	12/5
0	x_2	1	0	1	0	0	1	
	Z_j	4	0	4	0	0	4	
	$Z_j - C_j$		–1↑	0	0	0	4	

C_B	B	x_B	x_1	x_2	S_1	S_2	S_3
		C_j	1	4	0	0	0
1	x_1	3/2	1	0	1/2	0	–2
0	S_2	9/2	0	0	–5/2	1	7
4	x_2	1	0	1	0	0	1
	Z_j	11/2	1	4	1/2	0	2
	$Z_j - C_j$		0	0	1/2	0	2

Since all $Z_j - C_j \geq 0$, the solution is optimum, given by $x_1 = 3/2$

$x_2 = 1$, and Max $Z = 11/2$

Since $x_1 = 3/2$ is not an integer, this sub-problem is branched again.

Sub-Problem (2)

Max $Z = x_1 + 4x_2$

Subject to,

$$2x_1 + 4x_2 \leq 7$$
$$5x_1 + 3x_2 \leq 15$$
$$x_2 \geq 2$$
$$x_1, x_2 \geq 0.$$

C_B	B	x_B	x_1	x_2	S_1	S_2	S_3	A_1	Min x_B/x_2
		C_j	1	4	0	0	0	$-M$	
←0	S_1	7	2	(4)	1	0	0	0	7/4
0	S_2	15	5	3	0	1	0	0	15/3
$-M$	A_1	2	0	1	0	0	–1	1	2/1
	Z_j	$-2M$	0	$-M$	0	0	M	$-M$	
	$Z_j - C_j$		–1	$-M - 4$	0	0	M	0	
4	x_2	7/4	1/2	1	1/4	0	0	0	
0	S_2	39/4	7/2	0	–3/4	1	0	0	
$-M$	A_1	1/4	–1/2	0	–1/4	0	–1	1	
	Z_j	$7 - \frac{5M}{4}$	$2 + \frac{M}{2}$	4	$1 + \frac{M}{4}$	0	M	$-M$	
	$Z_j - C_j$		$\frac{M}{2} + 1$	0	$\frac{M}{4} + 1$	0	M	0	

Since all $Z_j - C_j \geq 0$, but an artificial variable A_1 is in the basis at positive level, there exists no feasible solution. Hence, this sub-problem is dropped.

In sub-problem (1) Since, $x_1 = 3/2$

we have, $1 \leq x_1 \leq 2$

$= x_1 \leq 1, \quad x_1 \geq 2$

Applying these two conditions separately in the sub-problem (1), we get two sub-problems.

Sub-problem (3)

Max $Z = x_1 + 4x_2$

Subject to, $2x_1 + 4x_2 \leq 7$

$5x_1 + 3x_2 \leq 15$

$x_2 \leq 1$

$x_1 \leq 1$

$x_1, x_2 \geq 0.$

Sub-problem (4)

Max $Z = x_1 + x_2$

Subject to, $2x_1 + 4x_2 \leq 7$

$5x_1 + 3x_2 \leq 15$

$x_2 \leq 1$

$x_1 \geq 2$

$x_1, x_2 \geq 0$

Sub-Problem (3)

Since all $Z_j - C_j \geq 0$, an optimum solution is obtained. It is given by, $x_1 = 1, x_2 = 1$ and Max $Z = 5$. Since this solution is integer-valued this sub-problem cannot be branched further. The lower bound of the objective function is 5.

		C_j	1	4	0	0	0	0	
C_B	B	x_B	x_1	x_2	S_1	S_2	S_3	S_4	Min $\frac{x_B}{x_2}$
0	S_1	7	2	4	1	0	0	0	7/4
0	S_2	15	5	3	0	1	0	0	15/3
←0	S_3	1	0	(1)	0	0	1	0	1/1
0	S_4	1	1	0	0	0	0	1	—
	Z_j	0	0	0	0	0	0	0	
	$Z_j - C_j$		–1	–4↑	0	0	0	0	Min $\frac{x_B}{x_1}$
0	S_1	3	2	0	1	0	–4	0	3/2
0	S_2	12	5	0	0	1	–3	0	12/5
4	x_2	1	0	1	0	0	1	0	—
←0	S_4	1	(1)	0	0	0	0	1	1/1
	Z_j	4	0	4	0	0	4	0	
	$Z_j - C_j$		–1↑	0	0	0	4	0	

		C_j	1	4	0	0	0	0
C_B	B	x_B	x_1	x_2	S_1	S_2	S_3	S_4
0	S_1	1	0	0	1	0	–4	–2
0	S_2	7	0	0	0	1	–3	–5
4	x_2	1	0	1	0	0	1	0
1	x_1	1	1	0	0	0	0	1
	Z_j	5	1	4	0	0	4	1
	$Z_j - C_j$		0	0	0	0	4	1

Sub-Problem (4)

Max $Z = x_1 + 4x_2$

Subject to,

$$2x_1 + 4x_2 \leq 7$$
$$5x_1 + 3x_2 \leq 15$$
$$x_2 \leq 1$$
$$x_1 \geq 2$$
$$x_1, x_2 \geq 0$$

Since all $Z_j - C_j \geq 0$, the solution is optimum and is given by,

$$x_1 = 2, \quad x_2 = 3/4$$

			C_j	1	4	0	0	0	0	$-M$	
	C_B	x_B	B	x_1	x_2	S_1	S_2	S_3	S_4	A_1	Min $\frac{x_B}{x_1}$
	0	7	S_1	2	4	1	0	0	0	0	7/2
	0	15	S_2	5	3	0	1	0	0	0	15/5
	0	1	S_3	0	1	0	0	1	0	0	—
←	$-M$	2	A_1	(1)	0	0	0	0	−1	1	2/1
		$-2M$	Z_j	$-M$	0	0	0	0	M	$-M$	
		—	$Z_j - C_j$	$-M-1$	−4	0	0	0	M	0	

		C_j	1	4	0	0	0	0	$-M$	
C_B	B	x_B	x_1	x_2	S_1	S_2	S_3	S_4	A_1	Min x_B/x_2
←0	S_1	3	0	(4)	1	0	0	2	−2	3/4
0	S_2	5	0	3	0	1	0	5	−5	5/3
0	S_3	1	0	1	0	0	1	0	0	1/1
1	x_1	2	1	0	0	0	0	−1	1	—
	Z_j	2	1	0	0	0	0	−2	1	
	$Z_j - C_j$		0	−4	0	0	0	−2	$1+M$	
4	x_2	3/4	0	1	1/4	0	0	1/2	—	
0	S_2	11/4	0	0	−3/4	1	0	7/2	—	
0	S_3	1/4	0	0	−1/4	0	1	−1/2	—	
1	x_1	2	1	0	0	0	0	−1	—	
	Z_j	5	1	4	1	0	0	1	—	
	$Z_j - C_j$		0	0	1	0	0	1	—	

Since $x_2 = 3/4$ is not an integer, this sub-problem is branched further.

In sub-problem (4) since $x_2 = 3/4$, $0 \leq x_2 \leq 1$

$$= x_2 \leq 0, \quad \text{or} \quad x_2 \geq 1$$

Applying these two conditions in the sub-problem (4)

We get two sub-problems.

Sub-problem (5)

Max $Z = x_1 + 4x_2$

Subject to, $2x + 4x_2 \leq 7$

$5x_1 + 3x_2 \leq 15$

Sub-problem (6)

Max $Z = x_1 + x_2$

Subject to, $2x_1 + 4x_2 \leq 7$

$5x_1 + 3x_2 \leq 15$

NOTES

$$x_2 \le 1 \qquad\qquad x_2 \le 1$$

$$x_1 \ge 2 \qquad\qquad x_1 \ge 2$$

$$x_2 \le 0 \qquad\qquad x_2 \ge 1$$

$$x_1, x_2 \ge 0 \qquad\qquad x_1, x_2 \ge 0.$$

Sub-Problem (5)

	C_B	B	x_B	x_1	x_2	S_1	S_2	S_3	S_4	A_1	S_5	Min x_B/x_1
		C_j		1	4	0	0	0	0	$-M$	0	
	0	S_1	7	2	4	1	0	0	0	0	0	7/2
	0	S_2	15	5	3	0	1	0	0	0	0	15/5
	0	S_3	1	0	1	0	0	1	0	0	0	—
←	$-M$	A_1	2	(1)	0	0	0	0	–1	1	0	2/1
	0	S_5	0	0	1	0	0	0	0	0	1	—
		Z_j	$-2M$	$-M$	0	0	0	0	M	$-M$	0	
		$Z_j - C_j$		$-M-1$	–4	0	0	0	M	0	0	Min x_B/x_2
	0	S_1	3	0	4	1	0	0	2	—	0	3/4
	0	S_2	5	0	3	0	1	0	(5)	—	0	5/3
	0	S_3	1	0	1	0	0	1	0	—	0	1/1
1	x_1	2	1	0	0	0	0	–1	—	0	—	
←	0	S_5	0	0	(1)	0	0	0	0	—	1	0/1
		Z_j	2	1	0	0	0	0	–2	—	0	
		$Z_j - C_j$		0	–4↑	0	0	0	–2	—	0	Min x_B/S_4
	0	S_1	3	0	0	1	0	0	2	—	0	3/2
←	0	S_2	5	0	0	0	1	0	5	—	0	1
	0	S_3	1	0	0	0	0	1	0	—	–1	—
	1	x_1	2	1	0	0	0	0	–1	—	0	—
	4	x_2	0	0	1	0	0	0	0	—	1	—
		Z_j	2	1	4	0	0	0	–1	—	0	
		$Z_j - C_j$		0	0	0	0	0	–1↑	—	0	
	0	S_1	1	0	0	1	–2/5	0	0	—	0	
	0	S_4	1	0	0	0	1/5	0	1	—	0	
	0	S_3	1	0	0	0	0	1	0	—	0	
	1	x_1	3	1	0	0	1/5	(1)	0	—	0	
	4	x_2	0	0	1	0	0	0	0	—	1	
		Z_j	3	1	4	0	3/5	0	0		4	
		$Z_j - C_j$		0	0	0	3/5	0	0	—	4	

Since all $Z_j - C_j \ge 0$, the solution is optimum and is given by $x_1 = 3, x_2 = 0$ and Max $Z = 3$. This sub-problem yields an optimum integer solution. Hence, this sub-problem is dropped.

Sub-Problem (6)

Max $Z = x_1 + 4x_2$

Subject to,

$$2x_1 + 4x_2 \le 7$$

$$5x_1 + 3x_2 \le 15$$

$$x_2 \le 1$$

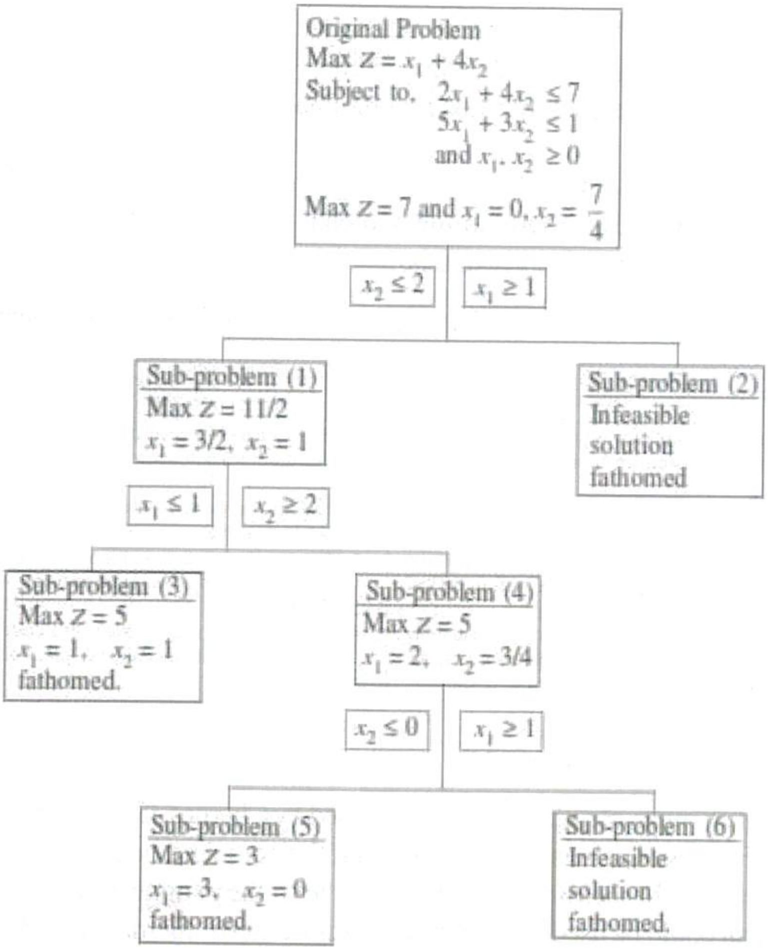
$$x_1 \ge 2$$

$$x_2 \ge 1$$

$$x_1, x_2 \ge 0$$

This sub-problem has no feasible solution. Hence, this sub-problem is also fathomed.

Original Problem



Among the available integer-valued solutions, the best integer solution is given by sub-problem (3).

∴ The optimum integer solution is,

$$\text{Max } Z = 5,\ x_1 = 1 \text{ and } x_2 = 1.$$

The best available integer optimal solution is,

$$\text{Max } Z = 5,\ x_1 = 1 \text{ and } x_2 = 1.$$

Check Your Progress

9. What do you mean by integer programming problem?
10. Define a pure integer programming problem.
11. Define a mixed integer programming problem.
12. Differentiate between pure and mixed IPP.
13. Mention a few applications of IPP.
14. Name the two methods used in solving IPP.

5.5 GOAL PROGRAMMING

Goal programming is an offshoot of multi-objective optimization. This is interconnected to the branch of multi-criteria decision analysis (MCDA), which is also known as multi-criteria decision making (MCDM). Thus, we can say that goal programming is an optimization programme, which is used to handle multiple and conflicting objective measures. These measures are given a goal or target value which they are supposed to be achieved. Often, it is thought to be an extension of the concept of linear programming, which is used to minimize the unwanted deviations from the set of target values.

Table 5.3 *Difference between goal programming and linear programming*

Goal Programming	*Linear Programming*
1. Has to handle multiple objective functions and attain optimization in multi-criteria decision making.	1. Has a single objective function to be optimized, like profit maximization and cost minimization.
2. Effective analysis for a decision-maker in a complex system of competing objectives.	2. Effective only when decision-maker is looking for single objective.
3. Goal programming problems use ordinal ranking of goals, which is decided on the basis of ordinal ranking of goals.	3. Linear programming problems are solved on the basis of cardinal value (exact amount).
4. Ordinal value, which calculates goal programming is based on their significance or impact on the organization.	4. Cardinal value, which is the basis of calculation is expressed in exact amount, i.e., profit or loss.

Concept of Goal Programming

Goal programming, as a concept, was first used by Charnes, Cooper and Ferguson in 1955. However, it was only in 1961 that the actual name appeared in a text introduced by Charnes and Cooper. In this text, they suggested the usage of a method that could be used for solving the multi-criteria dilemma faced due to the constraints of linear programming. Critical works on goal programming by Lee (1972) and Ignizio (1976) followed. This led to wide-scale usage of goal programming in planning, resource allocation, policy analysis and functional management issues. The first application of goal programming was done on an engineering application for design and placement of the antennas. This was during the second stage of Saturn V, which was used to launch the Apollo space capsule (this had landed the first men on the moon).

One of the landmark books on goal programming was written by Ijiri (1965) where he developed the concept of pre-emptive priority factors, assigning different priority levels to disproportionate goals and variant weights for the goals at the identical priority level. In goal programming, there is an achievement function that minimizes the deviations from the entrenched goal targets within a set of constraints. These are also known as slack variables (in the simplex algorithm of linear programming), and are used as dummy variables.

Terms: Objectives, Goals and Constraints

Objectives are referred to the optimization of the measure of performance of a decision, such as profit maximization or cost minimization.

Goals state a target value, i.e., the minimum acceptable level of performance of any decision taken by the decision-maker.

Constraints are similar to goals, in terms of their mathematical formulation. However, while goals are implied as the right-hand side value to achieve a certain target value, it is desirable for the constraints to achieve the right-hand side value.

Goal Programming Model Formulation

Single goal with Multiple Subgoals

An objective is the desired level of result by a decision-maker. This desired level of result (goal) may be underachieved, completely achieved, or overachieved within the given decision-making environment. The target level of any goal is determined by the relative managerial effort that is applied to an activity. In mathematical terms, one unit of applied effort towards activity x_j might contribute the amount a_{ij} towards the i^{th} goal. If this applied effort achieves its target value, the *ith* constraint would be denoted as:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i$$

However, one of the key features of goal programming is its flexibility and non-binding implications to mathematical interpretation. Thus, to allow underachievement or overachievement in the goal, we may denote:

d_i^- = negative deviation from i^{th} goal, i.e., below the target value.

d_i^+ = positive deviation from i^{th} goal, i.e., above the target value.

In the light of these notations, the i^{th} goal can be further written as:

$$\sum_{j=1}^{n} a_{ij} x_j + d_i^- - d_i^+ = b_i$$

$$\begin{bmatrix}\text{Value of the}\\ \text{objective}\end{bmatrix} + \begin{bmatrix}\text{Amount}\\ \text{below the}\\ \text{goal}\end{bmatrix} + \begin{bmatrix}\text{Amount}\\ \text{above the}\\ \text{goal}\end{bmatrix} = \text{Goal}$$

Here, *i* is equal to 1, 2,..., *m*.

It is important to note that it is not possible to achieve amount below the goal or above the goal simultaneously. In such a case, deviational variables of both the target values or goals $\left(d_i^- \text{ or } d_i^+\right)$ may be zero in the solution $\left(d_i^- \times d_i^+ = 0\right)$. Taking this at its optimal value, one may assume it to be of a positive value in the solution or must be kept at zero. The only significant point to be taken care of is that the goal deviational variables must be non-negative.

Note: Slack and surplus variables in the linear programming model are equivalent to the deviational variables in goal programming.

The surplus variable in linear programming and the deviational variable in goal programming, that is denoted as d_i^+, is done away from the objective function when there is a situation of overachievement. Similarly, in case of underachievement, d_i^- (known as slack variable or deviational variable) is removed from the objective function of goal programming. But, there is an exceptional situation in which both

d_i^- and d_i^+ are included in the objective function. This happens only when there is an exact attainment of the goal and it is ranked as per the pre-emptive priority factor.

NOTES

Example 5.26: A packaged food manufacture produces two kinds of products, chips and soda. The unit profit from a packet of chips is ₹ 80, and of a bottle of soda is ₹ 40. The goal of the plant manager is to earn a total profit of exactly ₹ 640 in the next week.

Model Formulation

We may interpret the profit goal in terms of subgoals, which are sales volume of chips and soda. Thereby, a goal programming model may be formulated as:

$$\text{Minimize, } z = d_i^- + d_i^+,$$

$$\text{subject to } 80x_1 + 40x_2 + d_i^- - d_i^+ = 640$$

$$x_1, x_2, d_i^-, d_i^+ \geq 0$$

where,

x_1 = number of packet of chips sold

x_2 = number of bottles of soda sold

d_i^- = underachievement of the profit goal of ₹ 640

d_i^+ = overachievement of the profit goal of ₹ 640

If the profit goal is not fully achieved, the slack in the profit goal will be expressed by d_i^- (negative deviational variable). Contrary to this situation, if the solution shows a profit in excess of ₹ 640, then both d_i^+ and d_i^- will be zero. Here it must be noted that d_i^- and d_i^+ are complementary to each other. So, if the profit goal of ₹ 640 is exactly achieved, both d_i^- and d_i^+ will be zero.

In the above example 5.26, there are an infinite number of combinations of x_1 and x_2 that would achieve the goal. The solution would be of any linear combination of x_1 and x_2 between the two points ($x_1 = 8, x_2 = 0$) or ($x_1 = 0, x_2 = 16$). This straight line is exactly the iso-profit function line when the total profit is ₹ 640.

Equally Ranked Multiple goals

This model given below can be extended to handle cases of multiple goals. Let us suppose that there are no model constraints.

Taking the same example as example 5.26:

Example 5.27: Let us consider that the package food manufacturer now desires to achieve a weekly profit as close as to ₹ 640 as possible. He wants to achieve sales volume for chips and soda close to six and four respectively. We can formulate this effort as a goal programming model.

Model Formulation

$$\text{Minimize, } z = d_1^- + d_2^+ + d_3^- + d_i^+$$

$$\text{Subject to, } 80x_1 + 40x_2 + d_i^- - d_i^+ = 640$$

$$x_1 + d_2^- = 6$$
$$x_2 + d_3^- = 4$$
$$x_1, x_2, d_1^-, d_1^+ d_2^-, d_3^- \geq 0$$

The above equation expresses the profit goal and the sales goals.

Here, d_2^- and d_3^- represent the underachievements of sales volume for chips and soda. It should be noted that d_2^+ and d_3^+ are not included in the second and third constraints, since the sales goal are given as the maximum sales volume. The solution to this problem can be formulated by a simple examination of the problem: if $x_2 = 6$ and $x_2 = 4$, then all targets will be completely achieved.

Therefore, $\quad d_1^- = d_2^- = d_3^- = d_1^+ = 0$

Ranking and Weighting of Unequal Multiple Goals

In general terms, goal programming model is a linear representation in which the optimum attainment of objectives is sought within the given decision environment. It is this decision environment which determines the basic component of the model, like constraints, decision variables and the objective function. Since multiple and conflicting goals are usually not of equal importance, negative or positive deviations are not added. To achieve the goals as per the pre-emptive priority factor, p_1, p_2... and so on are assigned to deviational variables in the formulation of the objective function to be minimized. The p_s does not assume any numerical value, so they are simply a convenient way of indicating that one goal is comparatively more important than another. The relationship between various priority factors is based on priority ranking, such as $p_1 >> p_2 >> ... p_k >> p_{k+1}$ ..., where >> 'denotes more important than'.

This further means, $p_j >> p_{j+1}$ ($j = 1, 2, ..., k$) where n is a large number. The priority ranking of any target value or goal cannot be improved by multiplying by n. Thus, it is important to note that the deviational variables at a similar priority level must have the same unit of measurability. This can be well illustrated in the following example:

Example 5.28: An office furniture manufacturer produces two types of products: desks and chairs. For manufacturing a chair or a desk, the manufacturer requires one hour of production capacity in the plant (maximum production capacity is 50 hours per week). However, due to limited sales capacity, the maximum number of desks and chairs which could be sold are six and eight per week, respectively. The gross margin from the sale of a desk is ₹ 90 and ₹ 60 for a chair.

The manufacturer desires to determine the number of units of each desk and chair, which should be produced per week in consideration of the following set of goals:

Goal 1: Available production capacity should be utilized as much as possible but should not exceed 50 hours per week.

Goal 2: Sales of both the products (desks and chairs) produced per week should be as much as possible.

Goal 3: Overtime should not exceed 20 per cent of the available time.

NOTES

Model Formulation

We can formulate this problem according to a goal programming model so that the manufacturer may achieve his goals.

Suppose x_1 and x_2 = number of units of desk and chair produced. The first goal pertains to production capacity attainment with a target set at 50 hours per week. This constraint can be expressed as:

$$x_1 + x_2 + d_1^- + d_1^+ = 50$$

here, d_1^- = underutilization of production capacity

d_1^+ = overutilization of production capacity.

If this goal is not achieved, then d_1^- would take on a positive value and d_1^+ would be zero.

The second goal pertains to maximization of sales volume with a target of 6 units of desks and 8 units of chairs per week. The sales constraints can be expressed as:

$$x_1 + d_2^- = 6 \qquad x_2 + d_3^- = 8$$

Thus, it is important to note that the sales goals are the maximum possible sales volume, d_2^+ and d_3^+ will not appear as these constraints. So, here the overachievement of sales goals is ruled out.

The third goal looks for the minimization of overtime hours as minimum as possible. The constraint is denoted as:

$$d_1^+ + d_4^- - d_4^+ = 0.2\ (50) = 10$$

here,

d_4^- = overtime less than 20 per cent of goal constraint

d_4^+ = overtime more than 20 per cent of goal constraint

d_1^+ = overtime beyond 50 hours

Now, we an formulate a model to express the above given problem as a goal programming model.

Minimize (total deviation) $Z = d_1^+ + d_2^- + d_3^- + d_4^+$ subject to the above mentioned constraints,

(*i*) Production capacity constraint

$$x_1 + x_2 + d_1^- - d_1^+ = 50$$

(*ii*) Sales constraints

$$x_1 + d_2^- = 6$$

$$x_2 + d_3^- = 8$$

(*iii*) Overtime constraint

$$d_1^+ + d_4^- - d_4^+ = 10, \text{ and}$$

$$x_1, x_2, d_1^-, d_2^-, d_3^-, d_4^-, d_4^+ \geq 0.$$

Goal programming algorithm

The procedure, which is also known as algorithm, is used to formulate a goal programming model is summarized below:

1. Identification of the goals and constraints based on the availability of resources or constraints that may act as hindrances in way of achievement of the targets.
2. Prioritize the significance of each goal in such a way that goals with priority level p_1 are most important, followed by p_2 (next most important) and so on.
3. Define the decision variables.
4. Formulate the constraints in the same manner as linear programming model.
5. Develop an equation constraint-wise by adding deviational variables $\left(d_i^- + d_i^+\right)$. These variables indicate the possible deviations (below or above the target value or goal value).
6. Write the objective function in terms of minimizing a prioritized function of the deviational variables.

Goal programming approach for solving decision-making problems can be divided as:

(*i*) General goal programming model

(*ii*) Modified simplex method of goal programming

(*iii*) Alternative simplex method for goal programming

General Model

Assuming m goals, the general goal programming model may be stated as:

Minimize, $$Z = \sum_{i=1}^{m} w_i P_i \left(d_i^- + d_i^+\right)$$

This is subject to the linear constraints, and can be expressed as:

$$\sum_{j=1}^{m} a_{ij} x_j + d_i^- - d_i^+ = b_i; \; i = 1, 2, \ldots, m$$

and, $$x_j, d_i^-, d_i^+ \geq 0, \text{ for all } i \text{ and } j$$

$$d_i^- \times d_i^+ = 0$$

where, Z is the sum of the deviations from all desired goals. The w_i is the non-negative constants used for representing the relative weight to be assigned or allocated to the deviational variables d_i^-, d_i^+ within a certain priority level. The p_i are the priority levels that are assigned to each relevant deviational variables according to the priority rankings, such as $p_1 > p_2, \ldots, > p_n$. The a_{ij} are constraints that are attached to each variable (utilized and crucial for decision making), and b_i are the right handed side target values (goals) of each constraint. Keeping these in view, two goal programming models may be formulated as follows:

(*i*) system constraints (indirectly related to goals)

(*ii*) goal constraints (directly related to target values)

Example 5.29: A carpenter produces two products, desks and chairs. Each product must be processed through two departments. Department *A* delivers 30 hours of production capacity per day, and department *B* delivers 60 hours. Each unit of a desk requires 2 hours in department *A* and 6 hours in department *B*. Each unit of chair requires 3 hours in department *A* and 4 hours in department *B*. Management has prioritized the following target values and expects to achieve the following daily product mix:

p_1 : Minimize the underachievement of total production of desks and chairs for 10 units.

p_2 : Minimize the underachievement of production for 7 units of chairs.

p_3 : Minimize the underachievement of production for 8 units of desks.

We can formulate this problem as per the goal programming model and after that look to solve it using the graphical method.

Model Formulation: Let us suppose,

x_1 and x_2 = number of units of products, *i.e.*, desks and chairs manufactured.

d_i^- and d_i^+ = underachievement and overachivement associated with goal *i*, respectively.

Thereby, we have the goal programming model stated as follows:

Minimize,

$$Z = p_1 d_1^- + p_2 d_3^- + p_3 d_2^-$$

As this is subject to the constraints, we may say—

$$2x_1 + 3x_2 \leq 30$$
$$6x_1 + 4x_2 \leq 60$$
$$x_1 + x_2 + d_1^- - d_1^+ = 10$$
$$x_1 + d_2^- - d_2^+ = 8$$
$$x_1 + d_3^- - d_3^+ = 7$$
$$x_1, x_2, d_i^-, d_i^+ \leq 0, \text{ for all } i$$

The Algorithm

In order to solve and formulate example 5.29, we will use the graphical solution method. This requires algorithmic calculations. You may follow the given points:

(*i*) Graph all the system constraints and identify all the feasible solutions space. However, if there are no existing system constraints, then the feasible solution space is the first quadrant. Further, if there are no feasible solution space exists, there is no solution.

(*ii*) Graph the straight lines corresponding to the goal constraints, labelling the deviational variables.

(*iii*) Select the point or points, which are best suited to satisfy the highest priority goal. This may be done through the feasible solutions space that is identified in step *1*.

(*iv*) Consider the remaining goals in a sequential manner. Identify a point that would satisfy them to a great extent. Ensure that a goal set on the lower priority is not achieved at the cost of reducing the degree of achievement of higher priority target values.

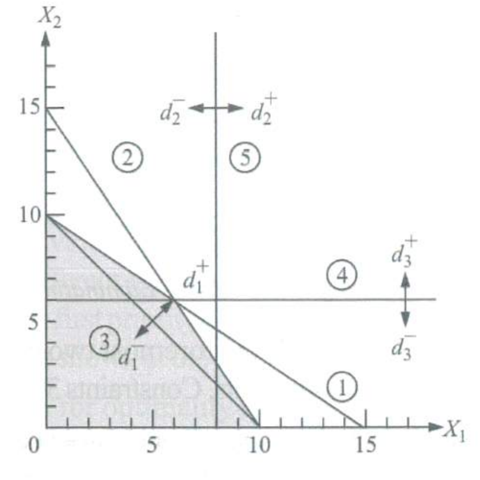


***Fig. 5.1(a)** Systems and Goal constraints*

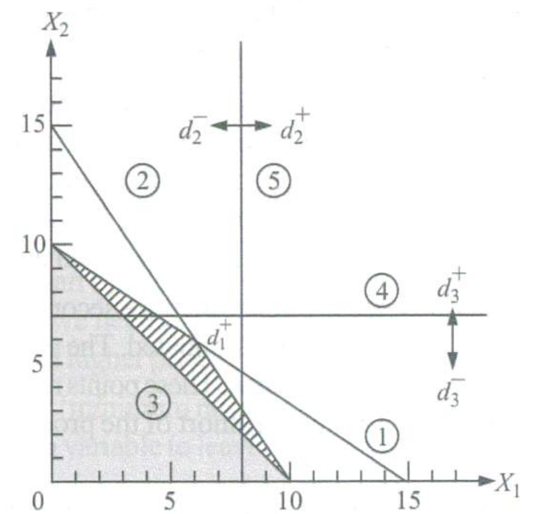


Fig. 5.1 (b) First goal Achieved—Eliminating d_1^- Area

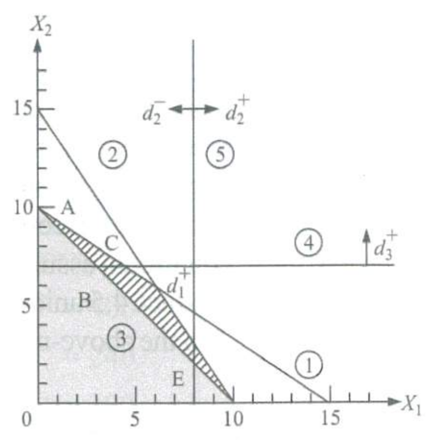


***Fig. 5.1(c)** Second goal Achieved—Eliminating* d_2^- *Area*

NOTES

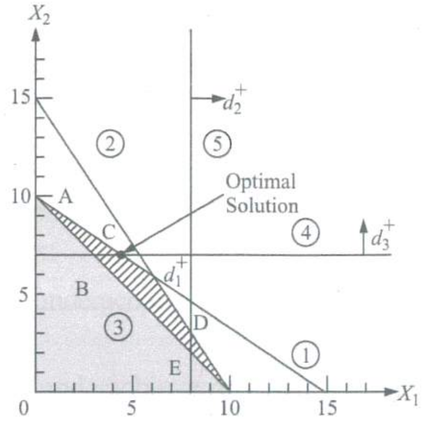


Fig. 5.1(d) Third goal Achieved—Eliminating d_3^- *Area*

Graphical Solution: In Figure 5.1, the foremost two constraints in this goal programming model are system constraints. Constraints 3 to 5 are goal constraints.

Figure 5.1(*a*) represents the solutions space that is associated with the two system constraints, while the lines are associated with the goal constraints with deviational variables (marked). The distance from the goal constraint line determine the value of deviational variable. This means, farther a point from a goal constraint, larger the value of corresponding deviational variable.

Figure 5.1(*b*) illustrates all possible points which have a positive value of d_1^- and have been eliminated. In this graph, the highest priority is to minimize the underachievement of the total production goal of ten units, represented as d_1^-. The lined area represent all combinations of products, *i.e.,* desks and chairs, which can be produced and exceed the production goal, *i.e.,* ten units.

In **Figure 5.1(*c*)** due significance is given to achieve second priority goal, where all points with positive values for d_3^- are eliminated. The points which lie below the second goal constraint line are eliminated. These points represent combinations of products (desks and chairs) which fall short of the production goal of seven units for chairs.

The third priority is established on the lined ABC point, which represents underachievement of the third goal (identify a point that would satisfy to a great extent).

Figure 5.1(*d*) wants to minimize d_2^- which represents the underachievement of the production goal of eight units of desks. This one point, known as **optimal solution,** occurs at corner point *C*, where d_3^- is contracted to as small as possible. At points *D* or *E*, positive values for d_3^- could be given and d_2^- would become zero. Here we see that point *C* occurs at the intersection point of constraints numbers 1 and 4. We can solve these equations, assuming $x_1 = 4.5$ and $x_2 = 7$. Thus, we find that the carpenter should produce 4.5 units of desks and seven units of chairs. Substituting $x_1 = 4.5$ and $x_2 = 7$ in the above-mentioned constraints, we may calculate:

(*a*) Department *A* utilized its maximum capacity of 30 hours.

(*b*) Department *B* has unutilized time, better known as slack, of 5 hours.

Thus, we can conclude that there is an overachievement of the total production goal equal to11.5 (4.5 + 17 – 10) units. The production goal of seven units of chairs has been completely achieved, however, there has been an underachievement in meeting the production goal for desks, with the carpenter making only 3.5 (8 – 4.5) units.

Modified Simplex Method of Goal Programming

The simplex method for a goal programming problem solution is somewhat similar to that of a linear programming problem. The salient features of this method for goal programming problem are:

(*i*) The z_j and $c_j - z_j$ values are calculated separately for each of the ranked goal $p_1, p_2, \ldots$. Different goals are measured in different units. On the basis of the priority, the first priority goal (p_1) is shown at the bottom and the least prioritized goal is shown at the top.

The criterion for optimality z_j or $c_j - z_j$ becomes a matrix $k \times n$ size. Here, k represents the number of pre-emptive priority levels, and n is the number of variables including decision and deviational variables.

(*ii*) Firstly, we need to examine $c_j - z_j$ values in the p_1 row. Optimal solution can be obtained, only if all $c_j - z_j \leq 0$ at the highest priority levels in the same column. If $c_j - z_j > 0$ at a certain priority level does not have any negative entry and in any similar column there is no higher unachieved priority levels, then the current optimal solution is not achieved.

(*iii*) The solution is optimal, if the target value of each goal in X_B- column is zero.

(*iv*) In order to determine the variable, which is needed to enter into the solution mix, we must start by examining $(c_j - z_j)$ row of highest priority (marked as p_1). After this, we need to select the largest negative value, else should more to the next higher priority (p_2) and select the largest negative value.

(*v*) To compute the 'minimum ratio' apply the usual procedure so that you are able to choose a variable to leave the current solution mix.

(*vi*) We need to ignore, if any negative value in the $(c_j - z_j)$ row has positive $(c_j - z_j)$ value under any lower priority rows. With the entry of this variable in the current solution mix, deviations from the highest priority goal would be increased.

Example 5.30: Using the modified simplex method to solve the following goal programming problem:

Minimize, $Z = p_1 d_1^- + p_2 \left(2d_2^- + d_3^-\right) + p_3 d_1^+$

Depending on the constraints,

$$x_1 + x_2 + d_1^- - d_1^+ = 400$$

$$x_1 + d_2^- = 240$$

$$x_2 + d_3^- = 300, \text{ and}$$

$$x_1, x_2, d_1^-, d_1^+, d_2^-, d_3^- \geq 0$$

Solution:

Initial Solution:

C_B	Variables in basics B	Solution values $b(= x_B)$	x_1	x_2	d_1^-	d_2^-	d_3^-	d_1^+	Minimum Ratio x_B/x_1
p_1	d_1^-	400	1	1	1	0	0	–1	400/1
$2p_2$	d_2^-	240	(1)	0	0	1	0	0	240/1
p_2	d_3^-	300	0	1	0	0	1	0	—
$c_j - z_j$	p_3	0	0	0	—	—	—	1	
	p_2	780	–2	–1	—	—	—	0	
	p_1	400	–1	–1	—	—	—	1	

The initial modified simplex table for the problem is given above. The initial table for the goal programming problem is formulated with the basic assumption that it is same as any linear programming problem. The c_j value in linear programming and goal programming are same, keeping the pre-emptive factors and difference weights corresponding.

In the above table, the criterion for optimality is 3 × 6 matrix as there are three priority levels and six variables in the model. Two are for decision and four are deviational. By using the standard simplex method for the computation of z - value, we can get the z - value, in goal programming as:

$$z = p_1 \times 400 + 2p_2 \times 240 + p_2 \times 300 = 400p_1 + 780p_2.$$

The values of the various priority levels, $p_1 = 400, p_2 = 780$ and $p_3 = 0$ in the X_B- column (represent the unachieved portion of each goal).

Now let us compute, the $c_j - z_j$ values as given in the table. As already confirmed, c_j values represent the priority factors which affect the deviational variables and z_j values represent the sum of the product of entries in C_B- column with columns of coefficient matrix. So, we can say that the $c_j - z_j$ values for each of the column is computed as follows:

$$c_1 - z_1 = 0 - (p_1 \times 1 + 2p_2 \times 1 + p_2 \times 0) = -p_1 - 2p_2$$
$$c_2 - z_2 = 0 - (p_1 \times 1 + 2p_2 \times 0 + p_2 \times 1) = -p_1 - p_2$$
$$c_6 - z_6 = p_3 - (p_1 \times -1) = p_3 + p_1$$

Alternative Simplex Method for Goal Programming

The alternative simplex method for goal programming can be explained through the following two steps:

Step 1: Formulation of initial solution trade

Step 2: Modify the initial solution

Now to explain it in detail,

Step 1: Formulation of Initial Solution Trade

The general formulation and calculation of an alternative simplex method for goal programming problem, can be done with some minimal alternations and modifications to the critical solution table. Firstly, the goal constraint needs to be reformulated in terms of their basic variables (d_i^+):

$$d_i^+ = -b_i + \sum_{j=1}^{n} a_{ij}\, x_j + d_i^- ; i \text{ here } i \text{ denotes } 1, 2, ..., m.$$

If any of the goal constraint does not possess a d_i^+ variable then, we may artificially induce a zero priority for formulating an initial solution table. The variables with zero value are non-basic variables.

Table 5.4 *Initial table for a generalized goal programming model*

	(1)	(2)	(3)	(4)	(5)
				$x_1, x_2, ..., x_n$	$d_1^-, d_2^-, ..., d_m^-$
(1)	Weighted priority	Z	$\sum_{j=1}^{m} [w_i - b_i]$	$0, 0, ..., 0$	$w_1, w_2, ..., w_m$
(2)	$w_1 p_1$	d_1^+	$-b_1$	$a_{11}, a_{12}, ..., a_{1n}$	1 0 ... 0
(3)	$w_2 p_2$ ...	d_2^+ ...	$-b_2$ ...	$a_{21}, a_{22}, ..., a_{2n}$	0 1 ... 0
	$w_m p_m$	d_m^+	$-b_m$	⋮	⋮
				$a_{m1}, a_{m2}, ..., a_{mn}$	0 0 ... 1

In the above Table 5.4,

(*i*) Row 1 lables the decision variables x_i and negative deviational variable d_i^-. Row 2 contains a value called 'total absolute deviation'.

(*ii*) The right-handed values, (b_i), are located in third column.

(*iii*) The decision-variable coefficients (a_{ij}) are located in the fourth column.

(*iv*) The fifth column has an identity matrix placed in it, which represents the inclusion of negative deviational variables (d_i^-). Owing to the fact that all deviational variables are not included, all problem formulations are equal to zero.

(*v*) Second column lists the artificial deviational variables alongwith the first column, which lists the appropriate priority factors p_i and weights w_i for each positive deviational variable.

Step 2: Modify the Initial Solution

First determine which variable is to be used to exit the solution basis. This can be accomplished by selecting the one with the highest ranked priority. We need to choose the variables with greatest mathematical weight, if two or more variables have the same priority ranking. However, as we look towards the calculation for an optimal solution, we need to choose a pivot row. This pivot row (positive coefficients) is divided into pivot column. Thus, the element found at the intersection of pivot row and pivot column is known as pivot element. Then, new alternative element corresponds by taking the reciprocal of pivot element. It can be formulated as:

$$\text{New Element} = \text{Old Element} - \frac{\text{Product of Two Corner Elements}}{\text{Pivot Element}}$$

NOTES

We can find the total absolute deviation by the following formula:

$$Z = \sum_{i=1}^{m} [w_i \cdot b_i]$$

Select the column with the smallest resulting ratio when the negative coefficients in the pivot row are divided into their positive elements in the second row charging the resulting sign, to determine which variable is to enter the solution basis.

Thus, only when the basic variables are all positive and objective functions have a negative sign, the solution is optimal.

Check Your Progress

15. What is goal programming?
16. Name the types of goal programming approach for solving decision-making problems.

5.6 ANSWERS TO 'CHECK YOUR PROGRESS'

1. A transportation problem deals with transportation of various quantities of a single homogeneous commodity initially stored at various origins to different destinations at the minimum cost.
2. The approaches used with transportation problems for determining the starting solution are as follows:
   (i) North West Corner Rule
   (ii) Least Cost Method (Matrix Minima)
   (iii) Vogel's Approximation Method
3. An optimal solution to a transportation problem is one which minimizes the total transportation cost.
4. The purpose of the MODI method is to get the optimal solution of a transportation problem.
5. The two conditions necessary for an alternate solution are as follows:
   (i) The slope of the objective function should be similar as constraint which forms the edge of the feasible region.
   (ii) The constraint must be an active constraint, that is, in the path of optimal progress of the objective function.
6. The one main difference between transportation problem and assignment problem is that in transportation problem the number of sources and destinations need not be equal. Hence, the cost matrix is not necessarily a square matrix. While in assignment problem the assignment is done on a one to one basis, the number of sources and destinations are equal. Hence, the cost matrix must be a square matrix.

7. Any assignment problem is said to be unbalanced if the cost matrix is not a square matrix, i.e., the number of rows and columns are not equal. To make it balanced, we add a dummy row or dummy column with all the entities as zero.
8. A network routing problem consists of finding an optimum route between two or more nodes in relation to total time, cost, or distance. Various constraints may exist, such as a prohibition on returning to a node already visited or a stipulation of passing through every node only once.
9. An LPP in which some or all the variables in the optimal solution are restricted to assume non-negative integer values is called an integer programming problem.
10. In a LPP, if all the variables in the optimal solution are restricted to assume non-negative integer value, then it is called a pure IPP.
11. In an LPP, if only some of the variables in the optimal solution are restricted to assume non-negative integer values, while the remaining variables are free to take any non-negative values, then it is called a mixed integer programming problem.
12. In a pure IPP all the variables in the optimal solution are restricted to assume non-negative integer values. Whereas in mixed IPP, only some of the variables in the optional solution are restricted to assume non-negative integer values.
13. Some applications of IPP are as follows:
    (i) In product mix problem
    (ii) Sequencing and routing decisions
    (iii) All allocation problems involving the allocation of goods, men and machine.
14. The two methods used in solving IPP are as follows:
    (i) Cutting methods (Gomany's cutting plane algorithm)
    (ii) Search method (Branch and Bound Technique)
15. Goal programming is an optimization programme, which is used to handle multiple and conflicting objective measures. These measures are given a goal or target value which they are supposed to be achieved.
16. Goal programming approach for solving decision-making problems can be divided as:
    (*i*) General goal programming model
    (*ii*) Modified simplex method of goal programming
    (*iii*) Alternative simplex method for goal programming

5.7 SUMMARY

- The transportation problem (TP) is one of the subclasses of LPP (Linear Programming Problem) in which the objective is to transport various quantities of a single homogeneous commodity that are initially stored at various origins

NOTES

to different destinations in such a way that the transportation cost is minimum.

- Any set of non-negative allocations ($X_{ij} > 0$) which satisfies the row and column sum (rim requirement) is called a feasible solution.
- A feasible solution is called a basic feasible solution if the number of non-negative allocations is equal to m + n – 1, where m is the number of rows and n the number of columns in a transportation table.
- Any feasible solution to a transportation problem containing m origins and n destinations is said to be non-degenerate, if it contains m + n – 1 occupied cells and each allocation is in independent positions.
- The allocations are said to be in independent positions if it is impossible to form a closed path. Closed path means by allowing horizontal and vertical lines and when all the corner cells are occupied.
- If a basic feasible solution contains less than m + n – 1 non-negative allocations, it is said to be degenerate basic feasible solution.
- An optimal solution is a feasible solution (not necessarily basic) which minimizes the total cost.
- The solution of a transportation problem (TP) can be obtained in two stages, namely initial solution and optimum solution.
- The cells in the transportation table can be classified into occupied cells and unoccupied cells. The allocated cells in the transportation table are called occupied cells and the empty cells in the transportation table are called unoccupied cells.
- Optimality test can be conducted to any initial basic feasible solution of a TP, provided such allocations has exactly m + n – 1 non-negative allocations, where m is the number of origins and n is the number of destinations. Also, these allocations must be in independent positions. To perform this optimality test, modified distribution method (MODI) is used.
- MODI method is applied in order to determine the optimum solution. One can determine a set of numbers u_i and v_j for each row and column, with $u_i + v_j = C_{ij}$ for each occupied cell. To start with, we give $u_2 = 0$ as the second row has the maximum number of allocation.
- An assignment problem is one of the fundamental combinatorial optimization problems and helps to find a maximum weight identical in nature in a weighted bipartite graph.
- The solution of an assignment problem can be arrived at using the Hungarian method.
- An assignment problem is balanced if the cost matrix is a square matrix; otherwise, it is termed as unbalanced.
- To convert an unbalanced assignment problem into a balanced problem, dummy rows or columns are added with all entries as 0s.

- A linear programming problem in which all or some of the decision variables are constrained to assume non-negative integer values is called an Integer Programming Problem (IPP).
- In a linear programming problem, if all variables are required to take integral values then it is called the Pure (all) Integer Programming Problem (Pure IPP).
- If only some of the variables in the optimal solution of a LPP are restricted to assume non-negative integer values, while the remaining variables are free to take any non-negative values, then it is called a Mixed Integer Programming Problem (Mixed IPP).
- If all the variables in the optimal solution are allowed to take values 0 or 1, then the problem is called the 0–1 Programming Problem or Standard Discrete Programming Problem.
- In LPP, all the decision variables were allowed to take any non-negative real values as it is quite possible and appropriate to have fractional values in many situations. There are several frequently occurring circumstances in business and industry that lead to planning models involving integer-valued variables.
- Integer programming is applied in business and industry. All assignment and transportation problems are integer programming problems, as in the assignment and travelling salesmen problem, all the decision variables are either zero or one.
- There are two methods used to solve IPP, namely Gomory's Cutting Plane Method and Branch and Bound Method (Search Method).
- Cutting method is a systematic procedure of solving pure IPP was first developed by R.E. Gomory, in 1956, which he later used to deal with the more complicated case of mixed integer programming problem.
- Cutting Method consists of first solving the IPP as an ordinary LPP by ignoring the restriction of integer values and then introducing a new constraint to the problem such that the new set of feasible solution includes all the original feasible integer solutions, but does not include the optimum non-integer solution initially found. This new constraint is called 'Fractional cut' or 'Gomorian constraint'.
- Search method is an enumeration method in which all feasible integer points are enumerated. The widely used search method is the Branch and Bound method. It was developed in 1960, by A.H. Land and A.G. Doig. This method is applicable to both pure and mixed IPP. It first divides the feasible region into smaller subsets that eliminate parts containing no feasible integer solution.
- In mixed IPP only some of the variables are restricted to integer values, while the other variables may take integer or other real values.
- Branch and Bound method is applicable to both, pure as well as mixed IPP. Sometimes a few or all the variables of an IPP are constrained by

their upper or lower bounds. The most general method for the solution of such constrained optimization problems is called 'Branch and Bound method'.

- If any sub-problem yields an optimum integer solution, it is not further branched. But if any sub-problem yields a non-integer solution, it is further branched into two sub-problems.
- The integer-valued solution among all the sub-problems, which gives the most optimal value of the objective function is then selected as the optimum solution.
- Goal programming is an optimization programme, which is used to handle multiple and conflicting objective measures. These measures are given a goal or target value which they are supposed to be achieved.
- The target level of any goal is determined by the relative managerial effort that is applied to an activity.
- The simplex method for a goal programming problem solution is somewhat similar to that of a linear programming problem.

5.8 KEY TERMS

- **Unbalanced Transportation:** The given TP is said to be unbalanced if the total supply is not equal to the total demand.
- **Vogel's Approximation Method (VAM):** It refers to iterative procedure calculated to find out the initial feasible solution of the transportation problem.
- **MODI Method:** The Modified Distribution Method or MODI is an efficient method of checking the optimality of the initial feasible solution.
- **Assignment problem:** It helps to find a maximum weight identical in nature in a weighted bipartite graph.
- **Unbalanced assignment problem:** Any assignment problem is said to be unbalanced if the cost matrix is not a square matrix.
- **Square matrix:** In mathematics, a square matrix is a matrix with the same number of rows and columns.
- **Integer Programming Problem (IPP):** It refers to a mathematical optimization or feasibility program in which some or all of the variables are restricted to be integers.
- **Cutting-plane method:** It refers to any of a variety of optimization methods that iteratively refine a feasible set or objective function by means of linear inequalities, termed cuts.
- **Branch and Bound Method:** It refers a systematic method for solving optimization problems.

5.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define feasible, basic, non-degenerate solutions of a transportation problem.
2. What are the numbers of non-basic variables for 4 rows and 5 columns?
3. While dealing with North West Corner rule, when does one move to the next cell in next column?
4. What is the coefficient of *Xij* of constraints in a transportation problem?
5. When does degeneracy occur in an $m \times n$ transportation problem?
6. State the differences between a transportation problem and an assignment problem.
7. State a mathematical formulation of the assignment problem.
8. How can an objective function in the assignment problem be maximised?
9. What are the applications of integer programming?
10. What is Standard Discrete Programming Problem?
11. Write a short informative note on mixed integer programming problem.
12. What is branch and bound technique?
13. What is the meaning of 'satisficing', and reasons for using the term in conjunction with goal programming?
14. Mention the importance of ranking goals in goal programming. What impact does it have on the solution of the problem?
15. Differentiate between goal programming and linear programming.

Long-Answer Questions

1. What do you understand by transportation model? Explain.
2. Describe the following with examples:
   (i) North West Corner Rule
   (ii) Least Cost Method
   (iii) Vogel's Approximation Method
3. Explain degeneracy in a transportation problem. Describe a method to resolve it.
4. What do you mean by an unbalanced transportation problem? Explain the process of converting an unbalanced transportation problem into a balanced one.
5. Illustrate the mathematical formulation of a transportation problem.
6. Explain an algorithm to solve a transportation problem.
7. Discuss the assignment problem with a suitable example.
8. Describe the algorithm for the solution of the assignment problem.

NOTES

9. Explain the nature of i in travelling salesman problem and give its mathematical formulation.

10. Solve the following assignment problems.

(a)

	A	B	C	D
I	1	4	6	3
II	9	7	10	9
III	4	5	11	7
IV	8	7	8	5

(b)

Men

Job	A	B	C	D
I	10	25	15	20
II	15	30	5	15
III	35	20	12	24
IV	17	25	24	20

11. **Men**

Tasks	A	B	C	D	E
I	1	3	2	8	8
II	2	4	3	1	5
III	5	6	3	4	6
IV	3	1	4	2	2
V	1	5	6	5	4

12. There are five jobs to be assigned, one each to 5 machines and the associated cost matrix is as follows.

Machine

Job	1	2	3	4	5
A	11	17	8	16	20
B	9	7	12	6	15
C	13	16	15	12	16
D	21	24	17	28	26
E	14	10	12	11	15

13. A salesman has to visit five cities A, B, C, D and E. The distance (in hundred miles) between the five cities is as follows.

To

From	A	B	C	D	E
A	-	7	6	8	4
B	7	-	8	5	6
C	6	8	-	9	7
D	8	5	9	-	8
E	4	6	7	8	-

If the salesman starts from city A and has to come back to his starting point, which route should he select so that the total distance travelled is minimum?

14. Determine the optimum assignment schedule for the following assignment problem. The cost matrix is given below.

Machine

Job	1	2	3	4	5	6
A	11	17	8	16	20	15
B	9	7	12	6	15	13
C	13	16	15	12	16	8
D	21	24	17	28	2	15
E	14	10	12	11	15	6

If the job C cannot be assigned to machine 6, will the optimum solution change?

15. Describe Gomarian constraint and its geometrical interpretation.
16. Why not round off the optimum values instead of resorting to integer programming? Explain.
17. Discuss the importance of IPP.
18. If you were the owner of company *A* and using goal programming to help in your decision making, what would be your goals? Please mention the kind of constraints which you would like to include in your model?
19. Mr Shah is the president and owner of Shah Prints, a company which prints two types of newspapers. The demand for newspaper *A* is upto 600 papers per week, demand for newspaper *B* is limited to 400 papers per week. Shah prints have a weekly operating power of 1500 hours. Newspaper *A* takes 2 hours to produce and newspaper *B* takes 3 hours. Each newspaper sold yields a profit of ' 10, and profit for a larger subscription model is ' 20.

    Mr. Shah has listed down the following goals in order of significance:

    (*i*) Attain a profit of ₹ 15,000 approximately each week.

    (*ii*) Ensure the firm's production capacity is not underutilized.

    (*iii*) Sell maximum number of newspaper *A* and newspaper *B*, as demand indicates.

    Explain with illustrations as a goal programming problem.
20. Miss Margaret Hall, Principal of Carmel Convent for girls, is concerned about 20 students taking a training programme and how they are spending their leisure time. Miss Hall has recognized the total needed hours per week is 170. She charted out a time-table as follows:

    $X1$ = number of hours of sleep needed per week

    $X2$ = number of hours for personal work

    $X3$ = number of class and study hours

    $X4$ = number of hours for social time.

    She assumes that the students must study 40 hours a week to have sufficient time to absorb subject-matter. This is the principal's major goal. Miss Margaret feels that students need at most 7 hours of sleep per night and she marked this goal at number 2. Further, she believes that goal 3 is to provide a minimum 20 hours of social time per week.

    (*a*) Formulate this as a goal programming problem.

    (*b*) Solve the problem using computer software.
21. Discuss the significance of goal programming in multiple choices for decision-making process.

5.10 FURTHER READING

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House.

Gupta, S. C. 2006. *Fundamentals of Statistics*. New Delhi: Himalaya Publishing House.

Gupta, S. P., 2005. *Statistical Methods*. New Delhi: Sultan Chand and Sons.

Hooda, R. P. 2002. *Statistics for Business and Economics*. New Delhi: Macmillan India.

Kothari, C. R., 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Monga, G. S. 2000. *Mathematics and Statistics for Economics*. New Delhi: Vikas Publishing House

Gupta, S.P. 2006. *Statistical Methods*. New Delhi: S. Chand & Co. Ltd.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Gupta, S.C. and V.K. Kapoor. *Fundamentals of Mathematical Statistics*. New Delhi: Sultan Chand & Sons.

Sharma, A. 2009. *Operations Research.* New Delhi: Himalaya Publishing House.

Singh, J. K. 2009. *Business Mathematics*. New Delhi: Himalaya Publishing House.

Kalavathy, S. 2010. *Operations Research with C Programmes,* 3rd edition. New Delhi: Vikas Publishing House Pvt. Ltd.

Kothari, C. R. 1992. *An Introduction to Operational Research.* New Delhi: Vikas Publishing House Pvt. Ltd.

MBA, Second Year
Paper - II

QUANTITATIVE TECHNIQUES FOR MANAGERS

Madhya Pradesh Bhoj (Open) University, Bhopal
(Established under an Act of State Assembly in 1991)

मध्यप्रदेश भोज (मुक्त) विश्वविद्यालय, भोपाल

Kolar Road, Near Swarn Jayanti Park, Yashoda Vihar Colony,
Chuna Bhatti Bhopal, Madhya Pradesh-462016

ISBN 978-93-5453-619-9

9 789354 536199