DJC2E - BUSINESS MATHEMATICS AND STATISTICS

Unit I

Business statistics – meaning – definition – uses and limitations – collection of primary and secondary data – sampling methods

Unit II

Measurement of central tendency – mean, median, mode, geometric mean, harmonic mean – advantages and disadvantages and calculation – measures of dispersion – range, quartile deviation, mean deviation, standard deviation – advantage, disadvantage and calculation – Skewness – Karlpearlson and Bowley's Coefficient of Skewness

Unit III

Correlation – meaning – types – calculation of Karlpearlsons co-efficient of correlation – rank correlation (only individual observation). Regression – meaning – calculation of X value and Y value – time series – meaning – components – uses – calculations, moving average – seasonal index and trend value by straight line method

Unit IV

Analytical geometry – point of distance between two points – slope of straight line – business application – modeling by liner functions

Unit V

Algebra – theory of indices – zero and negative indices – (without fractional indices – laws of indices – matrics – concepts – proof) – multiplication – matric inverse – solving

Introduction

In the modern world of computers and information technology, the importance of statistics is very well recognized by all the disciplines. Statistics has originated as a science of statehood and found applications slowly and steadily in Agriculture, Economics, Commerce, Biology, Medicine, Industry, planning, education and so on. As on date, there is no other human walk of life, where statistics cannot be applied.

The word 'Statistics' and 'Statistical' are all derived from the Latin word Status, means a political state. The theory of statistics as a distinct branch of the scientific method is of comparatively recent growth. Research, particularly into the mathematical theory of statistics, is rapidly proceeding and fresh discoveries are being made all over the world.

Statistics is concerned with scientific methods for collecting, organizing, summarising, presenting and analyzing data as well as deriving valid conclusions and making reasonable decisions on the basis of this analysis. Statistics is concerned with the systematic collection of numerical data and its interpretation. The word 'statistic' is used to refer to

1. Numerical facts, such as the number of people living in particular area.
2. The study of ways of collecting, analyzing and interpreting the facts.

Definitions

Statistics is defined differently by different authors over a period of time. In the olden day's statistics was confined to only state affairs but in modern days it embraces almost every sphere of human activity. Therefore a number of old definitions, which was confined to narrow field of enquiry were replaced by more definitions, which are much more comprehensive and exhaustive. Secondly, statistics has been defined in two different ways – Statistical data and statistical methods. The following are some of the definitions of statistics as numerical data.

1. Statistics are the classified facts representing the conditions of people in a state. In particular, they are the facts, which can be stated in numbers or in tables of numbers or in any tabular or classified arrangement.

2. Statistics are measurements, enumerations or estimates of natural phenomenon usually systematically arranged, analyzed and presented as to exhibit important interrelationships among them.

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

Statistics may be called the science of counting in one of the departments due to Bowley, obviously, this is an incomplete definition as it takes into account only the aspect of collection and ignores other aspects such as analysis, presentation, and interpretation. Bowley gives another definition for statistics, which states 'statistics may be rightly called the scheme of averages'. This definition is also incomplete, as averages play an important role in understanding and comparing data and statistics provide more measures.

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

It is clear that the definition of statistics by Croxton and Cowden is the most scientific and realistic one. According to this definition, there are four stages:

1. Collection of Data: It is the first step and this is the foundation upon which the entire data set. Careful planning is essential before collecting the data. There are different methods of collection of data such as census, sampling, primary, secondary, etc., and the investigator should make use of the correct method.

2. Presentation of data: The mass data collected should be presented in a suitable, concise form for further analysis. The collected data may be presented in the form of tabular or diagrammatic or graphic form.

3. Analysis of data: The data presented should be carefully analysed for making inference from the presented data such as measures of central tendencies, dispersion, correlation, regression etc.,

4. Interpretation of data: The final step is drawing a conclusion from the data collected. A valid conclusion must be drawn on the basis of analysis. A high degree of skill and experience is necessary for the interpretation.

"Statistics may be defined as the aggregate of facts affected to a marked extent by a multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner, for a predetermined purpose and placed in relation to each other." - ***Horace Secrist***

Functions of Statistics

There are many functions of statistics. Let us consider the following five important functions.

Condensation:

Generally speaking the word 'to condense', we mean to reduce or to lessen. Condensation is mainly applied at embracing the understanding of a huge mass of data by providing only a few observations. If in a particular class in Chennai School, only marks in an examination are given, no purpose will be served. Instead, if we are given the average mark in that particular examination, definitely it serves the better purpose. Similarly, the range of marks is also another measure of the data. Thus, Statistical measures help to reduce the complexity of the data and consequently to understand any huge mass of data.

Comparison:

Classification and tabulation are the two methods that are used to condense the data. They help us to compare data collected from different sources. Grand totals, measures of central tendency measures of dispersion, graphs and diagrams, the coefficient of correlation etc provide ample scope for comparison. If we have one group of data, we can compare within itself. If the rice production (in Tonnes) in Tanjore district is known, then we can compare one region with another region within the district. Or if the rice production (in Tonnes) of two different districts within Tamilnadu is known, then also a comparative study can be made. As statistics is an aggregate of facts and figures, the comparison is always possible and in fact, comparison helps us to understand the data in a better way.

Forecasting:

By the word forecasting, we mean to predict or to estimate before hand. Given the data of the last ten years connected to rainfall of a particular district in Tamilnadu, it is possible to predict or forecast the rainfall for the near future. In business also forecasting plays a dominant role in connection with the production, sales, profits etc. The analysis of time series and regression analysis plays an important role in forecasting.

Estimation:

One of the main objectives of statistics is drawn inference about a population from the analysis of the sample drawn from that population. The four major branches of statistical inference are

1. Estimation theory
2. Tests of Hypothesis
3. Non Parametric tests
4. Sequential analysis

In estimation theory, we estimate the unknown value of the population parameter based on the sample observations. Suppose we are given a sample of heights of hundred students in a school, based on the heights of these 100 students, it is possible to estimate the average height of all students in that school.

Tests of Hypothesis:

A statistical hypothesis is some statement about the probability distribution, characterising a population on the basis of the information available from the sample observations. In the formulation and testing of hypothesis, statistical methods are extremely useful. Whether crop yield has increased because of the use of new fertilizer or whether the new medicine is effective in eliminating a particular disease are some examples of statements of hypothesis and these are tested by proper statistical tools.

Scope of Statistics

Statistics is not a mere device for collecting numerical data, but as a means of developing sound techniques for their handling, analysing and drawing valid inferences from them. Statistics is applied in every sphere of human activity – social as well as physical – like Biology, Commerce, Education, Planning, Business Management, Information Technology, etc. It is almost impossible to find a single department of human activity where statistics cannot be applied. We now discuss briefly the applications of statistics in other disciplines.

Statistics and Industry:

Statistics is widely used in many industries. In industries, control charts are widely used to maintain a certain quality level. In production engineering, to find whether the product is conforming to specifications or not, statistical tools, namely inspection plans, control charts, etc., are of extreme importance. In inspection plans, we have to resort to some kind of sampling – a very important aspect of Statistics.

Statistics and Commerce:

Statistics are the lifeblood of successful commerce. Any businessman cannot afford to either by under stocking or having an overstock of his goods. In the beginning, he estimates the demand for his goods and then takes steps to adjust with his output or purchases. Thus statistics is indispensable in business and commerce. As so many multinational companies have invaded into our Indian economy, the size and volume of business are increasing. On one side the stiff competition is increasing whereas on the other side the tastes are changing and new fashions are emerging. In this connection, market survey plays an important role to exhibit the present conditions and to forecast the likely changes in future.

Statistics and Agriculture:

Analysis of variance (ANOVA) is one of the statistical tools developed by Professor R.A. Fisher, plays a prominent role in agriculture experiments. In tests of significance based on small samples, it can be shown that statistics are adequate to test the significant difference between two sample means. In an analysis of variance, we are concerned with the testing of equality of several population means.

For an example, five fertilizers are applied to five plots each of wheat and the yield of wheat on each of the plots is given. In such a situation, we are interested in finding out whether the effect of these fertilizers on the yield is significantly different or not. In other words, whether the samples are drawn from the same normal population or not. The answer to this problem is provided by the technique of ANOVA and it is used to test the homogeneity of several population means.

Statistics and Economics:

Statistical methods are useful in measuring numerical changes in complex groups and interpreting collective phenomenon. Nowadays the uses of statistics are abundantly made in any economic study. Both in economic theory and practice, statistical methods play an important role.

Alfred Marshall said, “Statistics are the straw only which I like every other economist has to make the bricks”. It may also be noted that statistical data and techniques of statistical tools are immensely useful in solving many economic problems such as wages, prices, production, distribution of income and wealth and so on. Statistical tools like Index numbers, time series Analysis, Estimation theory, Testing Statistical Hypothesis are extensively used in economics.

Statistics and Education:

Statistics is widely used in education. Research has become a common feature in all branches of activities. Statistics is necessary for the formulation of policies to start a new course, consideration of facilities available for new courses etc. There are many people engaged in research work to test the past knowledge and evolve new knowledge. These are possible only through statistics.

Statistics and Planning:

Statistics is indispensable in planning. In the modern world, which can be termed as the "world of planning", almost all the organizations in the government are seeking the help of planning for efficient working, for the formulation of policy decisions and execution of the same. In order to achieve the above goals, the statistical data relating to production, consumption, demand, supply, prices, investments, income expenditure etc and various advanced statistical techniques for processing, analysing and interpreting such complex data are of importance. In India statistics play an important role in planning, commissioning both at the central and state government levels.

Statistics and Medicine:

In Medical sciences, statistical tools are widely used. In order to test the efficiency of a new drug or medicine, t - the test is used or to compare the efficiency of two drugs or two medicines, t - test for the two samples is used. More and more applications of statistics are at present used in the clinical investigation.

Statistics and Modern applications:

Recent developments in the fields of computer technology and information technology have enabled statistics to integrate their models and thus make statistics a part of decision-making procedures of many organizations. There are so many software packages available for solving design of experiments, forecasting simulation problems etc. SYSTAT, a software package offers more scientific and technical graphing options than any other desktop statistics package. SYSTAT supports all types of scientific and technical research in various diversified fields as follows

1. Archaeology: Evolution of skull dimensions

2. Epidemiology: Tuberculosis

3. Statistics: Theoretical distributions

4. Manufacturing: Quality improvement

5. Medical research: Clinical investigations.

6. Geology: Estimation of Uranium reserves from ground water

Limitations of statistics

Statistics with all its wide application in every sphere of human activity has its own limitations. Some of them are given below.

1. Statistics is not suitable for the study of the qualitative phenomenon: Since statistics is basically a science and deals with a set of numerical data, it is applicable to the study of only these subjects of enquiry, which can be expressed in terms of quantitative measurements. As a matter of fact, a qualitative phenomenon like honesty, poverty, beauty, intelligence etc, cannot be expressed numerically and any statistical analysis cannot be directly applied to this qualitative phenomenon. Nevertheless, statistical techniques may be applied indirectly by first reducing the qualitative expressions to accurate quantitative terms. For example, the intelligence of a group of students can be studied on the basis of their marks in a particular examination.

2. Statistics do not study individuals: Statistics does not give any specific importance to the individual items, in fact, it deals with an aggregate of objects. Individual items, when they are taken individually do not constitute any statistical data and do not serve any purpose for any statistical enquiry.

3. Statistical laws are not exact: It is well known that mathematical and physical sciences are exact. But statistical laws are not exact and statistical laws are only approximations. Statistical conclusions are not universally true. They are true only on an average.

4. Statistics table may be misused: Statistics must be used only by experts; otherwise, statistical methods are the most dangerous tools in the hands of the inexpert. The use of statistical tools by the inexperienced and untraced persons might lead to wrong conclusions. Statistics can be easily misused by quoting wrong figures of data. As King says aptly ‘statistics are like the clay of which one can make a God or Devil as one pleases’.

5. Statistics is only, one of the methods of studying a problem: statistical methods do not provide complete solution of the problems because problems are to be studied taking the background of the countries culture, philosophy or religion into consideration. Thus the statistical study should be supplemented by other evidence.

Questions

1. Explain the origin and growth of statistics
2. What are the uses of Statistics?
3. What do you understand by the term "Average" in statistics? Explain its significance in statistical work.
4. What are the limitations of statistics?
5. Explain about the applications of statistics in the field of commerce

Introduction

Everybody collects, interprets and uses information, much of it in numerical or statistical forms in day-to-day life. It is a common practice that people receive large quantities of information every day through conversations, televisions, computers, the radios, newspapers, posters, notices, and instructions. It is just because there is so much information available that people need to be able to absorb, select and reject it. In everyday life, in business and industry, certain statistical information is necessary and it is independent to know where to find it how to collect it. As consequences, everybody has to compare prices and quality before making any decision about what goods to buy. As employees of any firm, people want to compare their salaries and working conditions, promotion opportunities and so on. In time the firms on their part want to control costs and expand their profits.

One of the main functions of statistics is to provide information which will help on making decisions. Statistics provides the type of information by providing a description of the present, a profile of the past and an estimate of the future. The following are some of the objectives of collecting statistical information.

1. To describe the methods of collecting primary statistical information.
2. To consider the status involved in carrying out a survey.
3. To analyse the process involved in observation and interpreting.
4. To define and describe sampling.
5. To analyse the basis of sampling.
6. To describe a variety of sampling methods.

The statistical investigation is a comprehensive and requires systematic collection of data about some group of people or objects, describing and organizing the data, analyzing the data with the help of different statistical method, summarizing the analysis and using these results for making judgments, decisions and predictions. The validity and accuracy of final judgment are most crucial and depends heavily on how well the data was collected in the first place. The quality of data will greatly affect the conditions and hence at most importance must be given to this process and every possible precaution should be taken to ensure accuracy while collecting the data.

Categories of data

Any statistical data can be classified into two categories depending upon the sources utilized. These categories are,

1. Primary data
2. Secondary data

Primary data:

Primary data is the 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 a survey conducted by individuals or research institution or any organisation.

Example :

If a researcher is interested to know the impact of the noon meal scheme for the school children, he has to undertake a survey and collect data on the opinion of parents and children by asking relevant questions. Such a data collected for the purpose is called primary data.

The primary data can be collected by the following five methods.

1. Direct personal interviews.
2. Indirect Oral interviews.
3. Information from correspondents.
4. Mailed questionnaire method.
5. Schedules sent through enumerators.

1. Direct personal interviews:

The persons from whom information are collected are known as informants. The investigator personally meets them and asks questions to gather the necessary information. It is the suitable method for intensive rather than extensive field surveys. It suits best for intensive study of the limited field.

Merits:

1. People willingly supply information because they are approached personally. Hence, more response noticed in this method than in any other method.

2. The collected information is likely to be uniform and accurate. The investigator is there to clear the doubts of the informants.
3. Supplementary information on informant's personal aspects can be noted. Information on character and environment may help later to interpret some of the results.
4. Answers to questions about which the informant is likely to be sensitive can be gathered by this method.
5. The wordings in one or more questions can be altered to suit any informant. Explanations may be given in other languages also. Inconvenience and misinterpretations are thereby avoided.

Limitations:

1. It is very costly and time-consuming.
2. It is very difficult when the number of persons to be interviewed is large and the persons are spread over a wide area.
3. Personal prejudice and bias are greater under this method.

2. Indirect Oral Interviews:

Under this method the investigator contacts witnesses or neighbors or friends or some other third parties who are capable of supplying the necessary information. This method is preferred if the required information is on addiction or cause of fire or theft or murder etc., If a fire has broken out a certain place, the persons living in neighbourhood and witnesses are likely to give information on the cause of the fire. In some cases, police interrogated third parties who are supposed to have knowledge of a theft or a murder and get some clues. Enquiry committees appointed by governments generally adopt this method and get people' s views and all possible details of facts relating to the enquiry. This method is suitable whenever direct sources do not exist or cannot be relied upon or would be unwilling to part with the information.

The validity of the results depends on a few factors, such as the nature of the person whose evidence is being recorded, the ability of the interviewer to draw out information from the third parties by means of appropriate questions and cross examinations, and the number of persons interviewed. For the success of this method one person or one group alone should not be relied upon.

3. Information from correspondents:

The investigator appoints local agents or correspondents in different places and compiles the information sent by them. Information to Newspapers and some departments of Government come by this method. The advantage of this method is that it is cheap and appropriate for extensive investigations. But it may not ensure accurate results because the correspondents are likely to be negligent, prejudiced and biased. This method is adopted in those cases where information is to be collected periodically from a wide area for a long time.

4. Mailed questionnaire method:

Under this method, a list of questions is prepared and is sent to all the informants by post. The list of questions is technically called questionnaire. A covering letter accompanying the questionnaire explains the purpose of the investigation and the importance of correct information and request the informants to fill in the blank spaces provided and to return the form within a specified time. This method is appropriate in those cases where the informants are literates and are spread over a wide area.

Merits:

1. It is relatively cheap.
2. It is preferable when the informants are spread over the wide area.

Limitations:

1. The greatest limitation is that the informants should be literates who are able to understand and reply the questions.
2. It is possible that some of the persons who receive the questionnaires do not return them.
3. It is difficult to verify the correctness of the information furnished by the respondents.

With the view of minimizing non-respondents and collecting correct information, the questionnaire should be carefully drafted. There is no hard and fast rule. But the following general principles may be helpful in framing the questionnaire. A covering letter and a self-addressed and stamped envelope should accompany the questionnaire. The covering letter should politely point out the purpose of the survey and privilege of the respondent who is one among the few associated with the investigation. It should assure that the information would be kept confidential and would never be misused. It may promise a copy of the findings or free gifts or concessions etc.,

Characteristics of a good questionnaire

1. A number of questions should be minimum.
2. Questions should be in logical orders, moving from easy to more difficult questions.
3. Questions should be short and simple. Technical terms and vague expressions capable of different interpretations should be avoided.
4. Questions fetching YES or NO answers are preferable. There may be some multiple choice questions requiring lengthy answers are to be avoided.
5. Personal questions and questions which require memory power and calculations should also be avoided.
6. The question should enable cross check. Deliberate or unconscious mistakes can be detected to an extent.
7. Questions should be carefully framed so as to cover the entire scope of the survey.
8. The wording of the questions should be proper without hurting the feelings or arousing resentment.
9. As far as possible confidential information should not be sought.
10. Physical appearance should be attractive, sufficient space should be provided for answering each question.

5. Schedules sent through Enumerators:

Under this method, enumerator or interviewers take the schedules, meet the informants and filling their replies. Often a distinction is made between the schedule and a questionnaire. A schedule is filled by the interviewers in a face-to-face situation with the informant. A questionnaire is filled by the informant which he receives and returns by post. It is suitable for extensive surveys.

Merits:

1. It can be adopted even if the informants are illiterates.
2. Answers to questions of personal and pecuniary nature can be collected.
3. Non-response is minimum as enumerators go personally and contact the informants.
4. The information collected is reliable. The enumerators can be properly trained for the same.
5. It is most popular methods.

Limitations:

1. It is the costliest method.

2. Extensive training is to be given to the enumerators for collecting correct and uniform information.
3. Interviewing requires experience. Unskilled investigators are likely to fail in their work.

Before the actual survey, a pilot survey is conducted. The questionnaire/Schedule is pre-tested in a pilot survey. A few among the people from whom actual information is needed are asked to reply. If they misunderstand a question or find it difficult to answer or do not like its wordings etc., it is to be altered. Further, it is to be ensured that every question fetches the desired answer.

Merits and Demerits of primary data:

1. The collection of data by the method of personal survey is possible only if the area covered by the investigator is small. Collection of data by sending the enumerator is bound to be expensive. Care should be taken twice that the enumerator record corrects information provided by the informants.
2. Collection of primary data by framing schedules or distributing and collecting questionnaires by post is less expensive and can be completed in shorter time.
3. Suppose the questions are embarrassing or of complicated nature of the questions probe into personal affairs of individuals, then the schedules may not be filled with accurate and correct information and hence this method is unsuitable.
4. The information collected for primary data is more reliable than those collected from the secondary data.

Secondary Data:

Secondary data are those data which have been already collected and analysed by some earlier agency for its own use, and later the same data are used by a different agency. According to W.A.Neiswanger, ‘ A primary source is a publication in which the data are published by the same authority which gathered and analysed them. A secondary source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible’.

Sources of Secondary data:

In most of the studies the investigator finds it impracticable to collect first-hand information on all related issues and as such he makes use of the data collected by others. There is a vast amount of published information from which statistical studies may be made and fresh statistics are constantly in a state of production. The sources of secondary data can broadly be classified under two heads:

1. Published sources, and
2. Unpublished sources.

1. Published Sources:

The various sources of published data are:

1. ***Reports and official publications*** of

(i) International bodies such as the International Monetary Fund, International Finance Corporation, and United Nations Organisation.

(ii) Central and State Governments such as the Report of the Tandon Committee and Pay Commission.

2. ***Semi-official publication*** of various local bodies such as Municipal Corporations and District Boards.

3***. Private publications***-such as the publications of –

(i) Trade and professional bodies such as the Federation of Indian Chambers of Commerce and Institute of Chartered Accountants.

(ii) Financial and economic journals such as ‘Commerce’, ‘Capital’ and ‘Indian Finance’.

(iii) Annual reports of joint stock companies.

(iv) Publications brought out by research agencies, research scholars, etc.

It should be noted that the publications mentioned above vary with regard to the period of publication. Some are published at regular intervals (yearly, monthly, weekly etc.,) whereas others are ad hoc publications, i.e., with no regularity about periodicity of publications.

Note: A lot of secondary data is available on the internet. We can access it at any time for the further studies.

2. Unpublished Sources

All statistical material is not always published. There are various sources of unpublished data such as records maintained by various Government and private offices, studies made by research institutions, scholars, etc. Such sources can also be used where necessary

Precautions in the use of Secondary data

The following are some of the points that are to be considered in the use of secondary data

1. How the data has been collected and processed
2. The accuracy of the data
3. How far the data has been summarized
4. How comparable the data is with other tabulations
5. How to interpret the data, especially when figures collected for one purpose is used for another

Generally speaking, with secondary data, people have to compromise between what they want and what they are able to find.

Merits and Demerits of Secondary Data:

1. Secondary data is cheap to obtain. Many government publications are relatively cheap and libraries stock quantities of secondary data produced by the government, by companies and other organisations.
2. Large quantities of secondary data can be got through the internet.
3. Much of the secondary data available has been collected for many years and therefore it can be used to plot trends.
4. Secondary data is of value to:
    - The government – help in making decisions and planning future policy.
    - Business and industry – in areas such as marketing, and sales in order to appreciate the general economic and social conditions and to provide information on competitors.
    - Research organisations – by providing social, economic and industrial information.

Questions

1. Explain about secondary data
2. Explain the methods for collection primary and secondary data
3. Explain about primary data

Introduction

Sampling is very often used in our daily life. For example, while purchasing food grains from a shop we usually examine a handful from the bag to assess the quality of the commodity. A doctor examines a few drops of blood as a sample and draws a conclusion about the blood constitution of the whole body. Thus most of our investigations are based on samples.

Population

In a statistical enquiry, all the items, which fall within the purview of enquiry, are known as **Population** or **Universe**. In other words, the population is a complete set of all possible observations of the type which is to be investigated. A total number of students studying in a school or college, the total number of books in a library, the total number of houses in a village or town is some examples of the population.

Sometimes it is possible and practical to examine every person or item in the population we wish to describe. We call this a **Complete enumeration**, or **census**. We use **sampling** when it is not possible to measure every item in the population. Statisticians use the word population to refer not only to people but to all items that have been chosen for study.

Finite population and infinite population

A population is said to be finite if it consists of a finite number of units. A number of workers in a factory, production of articles in a particular day for a company are examples of a finite population. The total number of units in a population is called population size. A population is said to be infinite if it has an infinite number of units. For example the number of stars in the sky, the number of people seeing the Television programs etc.,

Census Method:

Information on population can be collected in two ways – census method and sample method. In census method, every element of the population is included in the investigation. For example, if we study the average annual income of the families of a particular village or area, and if there are 1000 families in that area, we must study the income of all 1000 families. In this method, no family is left out, as each family is a unit.

Merits and limitations of Census method:

Merits:

1. The data are collected from each and every item of the population
2. The results are more accurate and reliable because every item of the universe is required.
3. Intensive study is possible.
4. The data collected may be used for various surveys, analyses etc.

Limitations:

1. It requires a large number of enumerators and it is a costly method
2. It requires more money, labour, time energy etc.
3. It is not possible in some circumstances where the universe is infinite.

Sampling:

The theory of sampling has been developed recently but this is not new. In our everyday life, we have been using sampling theory as we have discussed in the introduction. In all those cases we believe that the samples give a correct idea about the population. Most of our decisions are based on the examination of a few items that are sample studies.

Sample:

Statisticians use the word **sample** to describe a portion chosen from the population. A finite subset of statistical individuals defined in a population is called a sample. The number of units in a sample is called the **sample size.**

Sampling unit:

The constituents of a population which are individuals to be sampled from the population and cannot be further subdivided for the purpose of the sampling at a time are called sampling units. For example, to know the average income per family, the head of the family is a sampling unit. To know the average yield of rice, each farm owner' s yield of rice is a sampling unit.

Sampling frame:

For adopting any sampling procedure it is essential to have a list identifying each sampling unit by a number. Such a list or map is called sampling frame. A list of voters, a list of house holders, a list of villages in a district, a list of farmers etc. are a few examples of the sampling frame.

Reasons for selecting a sample:

Sampling is inevitable in the following situations:

1. Complete enumerations are practically impossible when the population is infinite.
2. When the results are required in a short time.
3. When the area of the survey is wide.
4. When resources for the survey are limited particularly in respect of money and trained persons.
5. When the item or unit is destroyed under investigation.

Parameters and statistics:

We can describe samples and populations by using measures such as the mean, median, mode and standard deviation. When these terms describe the characteristics of a population, they are called **parameters.** When they describe the characteristics of a sample, they are called **statistics.** A parameter is a characteristic of a population and a statistic is a characteristic of a sample. Since samples are subsets of population statistics provide estimates of the parameters. That is, when the parameters are unknown, they are estimated from the values of the statistics. In general, we use Greek or capital letters for population parameters and lower case Roman letters to denote sample statistics. [N, μ, σ, are the standard symbols for the size, mean, S.D, of the population. N, x, s, are the standard symbol for the size, mean, s.d of sample respectively].

Principles of Sampling

Samples have to provide good estimates. The following principle tells us that the sample methods provide such good estimates.

1. The principle of statistical regularity:

A moderately large number of units chosen at random from a large group are almost sure on the average to possess the characteristics of the large group.

2. The principle of Inertia of large numbers:

Other things being equal, as the sample size increases, the results tend to be more accurate and reliable.

3. The principle of Validity:

This states that the sampling methods provide valid estimates of the population units (parameters).

4. The principle of Optimisation:

This principle takes into account the desirability of obtaining a sampling design which gives optimum results. This minimizes the risk or loss of the sampling design. The foremost purpose of sampling is to gather maximum information about the population under consideration at minimum cost, time and human power. This is best achieved when the sample contains all the properties of the population.

Sampling errors and nonsampling errors

The two types of errors in a sample survey are sampling errors and no – sampling errors.

1. Sampling errors:

Although a sample is a part of the population, it cannot be expected generally to supply full information about the population. So there may be in most cases difference between statistics and parameters. The discrepancy between a parameter and its estimate due to sampling process is known as **sampling error.**

2. Non-sampling errors:

In all surveys some errors may occur during collection of actual information. These errors are called Non-sampling errors.

Advantages and Limitation of Sampling:

There are many advantages of sampling methods over census method. They are as follows:

1. Sampling saves time and labour.
2. It results in reduction of cost in terms of money and manhour.
3. Sampling ends up with greater accuracy of results.
4. It has greater scope.
5. It has greater adaptability.
6. If the population is too large, or hypothetical or destroyable sampling is the only method to be used.

LIMITATIONS

The limitations of sampling are given below:

1. Sampling is to be done by qualified and experienced persons. Otherwise, the information will be unbelievable.
2. Sample method may give the extreme values sometimes instead of the mixed values.
3. There is the possibility of sampling errors. Census survey is free from sampling error.

Types of Sampling

The technique of selecting a sample is of fundamental importance in sampling theory and it depends upon the nature of investigation. The sampling procedures which are commonly used may be classified as

1. Probability sampling.
2. Non-probability sampling.
3. Mixed sampling.

Probability sampling (Random sampling):

A probability sample is one where the selection of units from the population is made according to known probabilities. (eg.) Simple random sample, probability proportional to sample size etc.

Non-Probability sampling:

It is the one where discretion is used to select 'representative' units from the population (or) to infer that a sample is 'representative' of the population. This method is called **judgement or purposive** sampling. This method is mainly used for opinion surveys; A common type of judgement sample used in surveys is quota sample. This method is not used in general because of prejudice and bias of the enumerator. However if the enumerator is experienced and expert, this method may yield valuable results. For example, in the market research survey of the performance of their new car, the sample was all new car purchasers.

Mixed Sampling:

Here samples are selected partly according to some probability and partly according to a fixed sampling rule; they are termed as mixed samples and the technique of selecting such samples is known as **mixed sampling**.

Methods of selection of samples:

Here we shall consider the following three methods:

1. Simple random sampling.

2. Stratified random sampling.

3. Systematic random sampling.

1. Simple random sampling:

A simple random sample from finite population is a sample selected such that each possible sample combination has equal probability of being chosen. It is also called unrestricted random sampling.

2. Simple random sampling without replacement:

In this method the population elements can enter the sample only once (ie) the units once selected is not returned to the population before the next draw.

3. Simple random sampling with replacement:

In this method the population units may enter the sample more than once. Simple random sampling may be with or without replacement.

Methods of selection of a simple random sampling:

The following are some methods of selection of a simple random sampling.

a) Lottery Method:

This is the most popular and simplest method. In this method all the items of the population are numbered on separate slips of paper of same size, shape and colour. They are folded and mixed up in a container. The required numbers of slips are selected at random for the desire sample size. For example, if we want to select 5 students, out of 50 students, then we must write their names or their roll numbers of all the 50 students on slips and mix them. Then we make a random selection of 5 students. This method is mostly used in lottery draws. If the universe is infinite this method is inapplicable.

b) Table of Random numbers:

As the lottery method cannot be used, when the population is infinite, the alternative method is that of using the table of random numbers. There are several standard tables of random numbers.

1. Tippett' s table

2. Fisher and Yates' table

3. Kendall and Smith' s table are the three tables among them.

A random number table is so constructed that all digits 0 to 9 appear independent of each other with equal frequency. If we have to select a sample from population of size N=100, then the numbers can be combined three by three to give the numbers from 001 to 100.

c) Random number selections using calculators or computers:

Random number can be generated through scientific calculator or computers. For each press of the key get a new random numbers. The ways of selection of sample is similar to that of using random number table.

Merits of using random numbers:

Merits:

1. Personal bias is eliminated as a selection depends solely on chance .
2. A random sample is in general a representative sample for a homogenous population.
3. There is no need for the thorough knowledge of the units of the population.
4. The accuracy of a sample can be tested by examining another sample from the same universe when the universe is unknown.
5. This method is also used in other methods of sampling.

Limitations:

1. Preparing lots or using random number tables is tedious when the population is large.
2. When there is large difference between the units of population, the simple random sampling may not be a representative sample.
3. The size of the sample required under this method is more than that required by stratified random sampling.
4. It is generally seen that the units of a simple random sample lie apart geographically. The cost and time of collection of data are more.

Stratified Random Sampling:

Of all the methods of sampling the procedure commonly used in surveys is stratified sampling. This technique is mainly used to reduce the population heterogeneity and to increase the efficiency of the estimates. Stratification means division into groups.In this method the population is divided into a number of subgroups or strata. The strata should be so formed that each stratum is homogeneous as far as possible. Then from each stratum a simple random sample may be selected and these are combined together to form the required sample from the population.

Types of Stratified Sampling:

There are two types of stratified sampling. They are **proportional** and **nonproportional**. In the proportional sampling equal and proportionate representation is given to subgroups or strata. If the number of items is large, the sample will have a higher size and vice versa.

The population size is denoted by N and the sample size is denoted by ' n' the sample size is allocated to each stratum in such a way that the sample fractions is a constant for each stratum. That is given by n/N = c. So in this method each stratum is represented according to its size.

In non-proportionate sample, equal representation is given to all the sub-strata regardless of their existence in the population.

Merits and limitations of stratified sampling:

Merits:

1. It is more representative.
2. It ensures greater accuracy.
3. It is easy to administer as the universe is sub - divided.
4. Greater geographical concentration reduces time and expenses.
5. When the original population is badly skewed, this method is appropriate.
6. For non – homogeneous population, it may field good results.

Limitations:

1. To divide the population into homogeneous strata, it requires more money, time and statistical experience which is a difficult one.
1. Improper stratification leads to bias, if the different strata overlap such a sample will not be a representative one.

Systematic Sampling:

This method is widely employed because of its ease and convenience. A frequently used method of sampling when a complete list of the population is available is **systematic sampling.** It is also called **Quasi-random sampling.**

Selection procedure

The whole sample selection is based on just a random start . The first unit is selected with the help of random numbers and the rest get selected automatically according to some pre designed pattern is known as **systematic sampling.** With systematic random sampling every Kth element in the frame is selected for the sample, with the starting point among the first K elements determined at random.

Merits :

1. This method is simple and convenient.
2. Time and work is reduced much.
3. If proper care is taken result will be accurate.
4. It can be used in infinite population.

Limitations:

1.Systematic sampling may not represent the whole population.

2.There is a chance of personal bias of the investigators.

Systematic sampling is preferably used when the information is to be collected from trees in a forest, house in blocks, entries in a register which are in a serial order etc.

Introduction

In the study of a population with respect to one in which we are interested we may get a large number of observations. It is not possible to grasp any idea about the characteristic when we look at all the observations. So it is better to get one number for one group. That number must be a good representative one for all the observations to give a clear picture of that characteristic. Such representative number can be a central value for all these observations. This central value is called a measure of central tendency or an average or a measure of locations.

There are five averages. Among them mean, median and mode are called simple averages and the other two averages geometric mean and harmonic mean are called special averages.

Definitions

The meaning of average is nicely given in the following definitions.

"A measure of central tendency is a typical value around which other figures congregate."

"An average stands for the whole group of which it forms a part yet represents the whole."

"One of the most widely used set of summary figures is known as measures of location."

Characteristics for a good or an ideal average

The following properties should possess for an ideal average.

1. It should be rigidly defined.
2. It should be easy to understand and compute.
3. It should be based on all items in the data.
4. Its definition shall be in the form of a mathematical formula.
5. It should be capable of further algebraic treatment.
6. It should have sampling stability.
7. It should be capable of being used in further statistical computations or processing.

Besides the above requisites, a good average should represent maximum characteristics of the data; its value should be nearest to the most items of the given series.

Arithmetic mean or mean

Arithmetic mean or simply the mean of a variable is defined as the sum of the observations divided by the number of observations. If the variable x assumes n values x_1, x_2 … x_n then the

mean, $\bar{X}$, is given by

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots X_n}{N}$$

Example 1:

Calculate the mean for 2, 4, 6, 8, 10

Solution:

$$\bar{X} = \frac{2+4+6+8+10}{5}$$

$$\bar{X} = \frac{30}{5}$$

$$\bar{X} = 6$$

Short cut method

Under this method an assumed or an arbitrary average (indicated by A) is used as the basis of calculation of deviations from individual values. The formula is

$$\bar{X} = A \pm \frac{\sum d}{n}$$

Where, A = the assumed mean or any value in x

d = the deviation of each value from the assumed mean

Example 2 :

A student' s marks in 5 subjects are 75, 68, 80, 92, 56. Find his average mark.

Solution :

X	d = X-A
75	7
A 68	0

80	12
92	24
56	– 12
N=5	**∑d=31**

$\bar{X} = A \pm \frac{\sum d}{N}$

$= 68 + \frac{31}{5}$

= 68+6.2

= 74.2

Grouped Data: ***Discrete Series***

Direct Method

In case of *discrete series*, frequency against each of the observations is multiplied by the value of the observation. The values, so obtained, are summed up and divided by the total number of frequencies. Symbolically,

$$\bar{X} = \frac{\sum fx}{N}$$

Where *x* = the mid-point of individual class

f = the frequency of individual class

N = the sum of the frequencies or total frequencies

Short-cut method:

$\bar{X} = A \pm \frac{\sum fd}{N}$

Where, A = any value in x; N = total frequency

Example 3:

Given the following frequency distribution, calculate the arithmetic mean

Marks	:	64	63	62	61	60	59
Number of Students	:	8	18	12	9	7	6

Solution:

X	F	Fx	d = X-A	fd
64	8	512	2	16
63	18	1134	1	18
62 A	12	744	0	0
61	9	549	–1	– 9
60	7	420	–2	–14
59	6	354	–3	– 18
	N =60	**∑fx=3713**		**∑fd=– 7**

Direct method

$\bar{X} = \frac{\sum fx}{N}$

$\bar{X} = \frac{3713}{60}$

$\bar{X} = 61.88$

Short-cut method

$\bar{X} = A \pm \frac{\sum fd}{N}$

$\bar{X} = 62 - \frac{7}{60}$

$\bar{X} = 62 – 0.12$

$\bar{X} = 61.88$

Continuous Series

Here, class intervals are given. The process of calculating arithmetic mean in case of continuous series is same as that of a discrete series. The only difference is that the mid-points of various class intervals are taken. You should note that class intervals may be exclusive or inclusive or of unequal size. Example of exclusive class interval is, say, 0–10, 10–20 and so on. Example of inclusive class interval is, say, 0–9, 10–19 and so on. Example of unequal class interval is, say, 0–20, 20–50 and so on. In all these cases, calculation of arithmetic mean is done in a similar way.

Direct method

$$\bar{X} = \frac{\sum fm}{N}$$

Short-cut method

$$\bar{X} = A \pm \frac{\sum fd}{N}$$

Step deviation method

$$\bar{X} = A \pm \frac{\sum fd}{N} \times C$$

Where

$$d = \frac{X-A}{C}$$

A = any value in x

N = total frequency

c = width of the class interval

Example 4 :

Following is the distribution of persons according to different income groups. Calculate arithmetic mean.

Income Rs.(100)	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Number of persons	6	8	10	12	7	4	3

Solution :

Income C.I	Number of Persons (f)	Mid M	$d = \frac{M - A}{c}$	Fd
0-10	6	5	-3	-18
10-20	8	15	-2	-16
20-30	10	25	-1	-10
30-40	12	35	0	0
40-50	7	45	1	7
50-60	4	55	2	8
60-70	3	65	3	9
	N=50			**∑fd= -20**

$\bar{X} = A \pm \frac{\sum fd}{N} \times c$

$\bar{X} = 35 - \frac{20}{50} \times 10$

$\bar{X} = 35 - \frac{200}{50}$

$\bar{X} = 35 - 4$

$\bar{X} = 31$

Merits and demerits of Arithmetic mean :

Merits:

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. If the number of items is sufficiently large, it is more accurate and more reliable.
4. It is a calculated value and is not based on its position in the series.
5. It is possible to calculate even if some of the details of the data are lacking.
6. Of all averages, it is affected least by fluctuations of sampling.
7. It provides a good basis for comparison.

Demerits:

1. It cannot be obtained by inspection nor located through a frequency graph.
2. It cannot be in the study of qualitative phenomena not capable of numerical measurement i.e. Intelligence, beauty, honesty etc.,
3. It can ignore any single item only at the risk of losing its accuracy.
4. It is affected very much by extreme values.
5. It cannot be calculated for open-end classes.
6. It may lead to fallacious conclusions, if the details of the data from which it is computed are not given.

Weighted Arithmetic mean

For calculating simple mean, we suppose that all the values or the sizes of items in the distribution have equal importance. But, in practical life this may not be so. In case some items are more important than others, a simple average computed is not representative of the distribution. Proper weightage has to be given to the various items. For example, to have an idea of the change in cost of living of a certain group of persons, the simple average of the prices of the commodities consumed by them will not do because all the commodities are not equally important, e.g rice, wheat and pulses are more important than tea, confectionery etc., It is the weighted arithmetic average which helps in finding out the average value of the series after giving proper weight to each group.

Definition:

The average whose component items are being multiplied by certain values known as "weights" and the aggregate of the multiplied results are being divided by the total sum of their "weight".

If $x_1, x_2 \ldots x_n$ be the values of a variable x with respective weights of $w_1, w_2 \ldots w_n$ assigned to them, then

$$\text{Weighted A.M} = \bar{X}_w = \frac{W_1 X_1 + W_2 X_2 + \cdots + W_i X_i}{W_1 + W_2 + \cdots W_n} = \frac{\sum W_i X_i}{\sum W_i}$$

Uses of the weighted mean:

Weighted arithmetic mean is used in:

a. Construction of index numbers.

b. Comparison of results of two or more universities where number of students differ.

c. Computation of standardized death and birth rates.

Example 5:

Calculate weighted average from the following data

Designation	**Monthly salary (in Rs)**	**Strength of the cadre**
Class 1 officers	1500	10
Class 2 officers	800	20
Subordinate staff	500	70
Clerical staff	250	100
Lower staff	100	150

Solution:

Designation	**Monthly salary, x**	**Strength of the cadre, w**	**wx**
Class 1 officer	1500	10	15,000
Class 2 officer	800	20	16,000
Subordinate staff	500	70	35,000
Clerical staff	250	100	25,000
Lower staff	100	150	15,000
		$\sum W$=350	**$\sum wx$= 1,06,000**

Weighted A.M= $\bar{X}_w = \frac{\sum wx}{\sum w}$

$$\bar{X}_w = \frac{106000}{350}$$

$$\bar{X}_w = 302.86$$

Harmonic mean (H.M)

Harmonic mean of a set of observations is defined as the reciprocal of the arithmetic average of the reciprocal of the given values. If x_1, x_2…..x_n are n observations,

$$H.M = \frac{n}{\sum_{i=1}^{n} f\left(\frac{1}{x_i}\right)}$$

For frequency distribution $H.M = \frac{N}{\sum_{i=1}^{n} f\left(\frac{1}{x_i}\right)}$

Example 6:

From the given data calculate H.M 5, 10,17,24,30

X	**$\frac{1}{x}$**
5	0.2000
10	0.1000
17	0.0588
24	0.0417
30	0.0333
Total	**0.4338**

$$H.M = \frac{n}{\sum(\frac{1}{x})}$$

$$H.M = \frac{5}{0.4338} = 11.526$$

Example 7:

The marks secured by some students of a class are given below. Calculate the harmonic mean.

Marks	20	21	22	23	24	25
Number of Students	4	2	7	1	3	1

Solution :

Marks (X)	**No. of students (f)**	$\frac{1}{x}$	$f(\frac{1}{x})$
20	4	0.0500	0.2000
21	2	0.0476	0.0952
22	7	0.0454	0.3178
23	1	0.0435	0.0435
24	3	0.0417	0.1251
25	1	0.0400	0.0400
	N=18		$\sum f\left(\frac{1}{x}\right)$ **=0.8216**

$$H.M = \frac{N}{\sum f\left(\frac{1}{x}\right)}$$

$$H.M = \frac{18}{0.8216} = 21.91$$

Merits of H.M :

1. It is rigidly defined.
2. It is defined on all observations.
3. It is amenable to further algebraic treatment.
4. It is the most suitable average when it is desired to give greater weight to smaller observations and less weight to the larger ones.

Demerits of H.M :

1. It is not easily understood.
2. It is difficult to compute.
3. It is only a summary figure and may not be the actual item in the series
4. It gives greater importance to small items and is therefore, useful only when small items have to be given greater weightage.

Geometric mean

The geometric mean of a series containing n observations is the n^{th} root of the product of the values. If x_1, x_2…, x_n are observations then

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

$$= (x_1 . x_2 \ldots x_3)^{\frac{1}{n}}$$

$$\text{Log } G.M = \frac{1}{n} \log (x_1 . x_2 \ldots x_n)$$

$$= \frac{1}{n} (\log x_1 + \log x_2 \ldots + \log x_n)$$

$$= \frac{\sum \log x_i}{n}$$

$$GM = \text{Antilog } \frac{\sum \log x_i}{n}$$

For grouped data

GM= Antilog $[\frac{\sum f \log x_i}{N}]$

Example 8:

Calculate the geometric mean of the following series of monthly income of a batch of families 180, 250, 490, 1400, 1050.

X	log x
180	2.2553
250	2.3979
490	2.6902
1400	3.1461
1050	3.0212
N=5	**∑log x=13.5107**

GM= Antilog $\frac{\sum log\ x_i}{n}$

= Antilog $\frac{13.5107}{5}$

= Antilog $\frac{13.5107}{5}$

= Antilog $2.7021 = 503.6$

Example 9:

Calculate the average income per head from the data given below. Use geometric mean.

Class of people	Number of families	Monthly income per head (Rs.)
Landlords	2	5000
Cultivators	100	400
Landless-labours	50	200
Money-lenders	4	3750
Office Assistants	6	3000
Shop keepers	8	750
Carpenters	6	600
Weavers	10	300

Solution :

Class of people	Annual income (Rs) X	Number of families (f)	Log x	f log x
Landlords	5000	2	3.6990	7.398
Cultivators	400	100	2.6021	260.210
Landless-labours	200	50	2.3010	115.050
Money-lenders	3750	4	3.5740	14.296

Office Assistants	3000	6	3.4771	20.863
Shop keepers	750	8	2.8751	23.2008
Carpenters	600	6	2.7782	16.669
Weavers	300	10	2.4771	24.771
		N=186		**482.257**

GM= Antilog $[\frac{\sum f \log x_i}{N}]$

GM= Antilog $[\frac{482.257}{186}]$

GM= Antilog [2.5928]

= Rs. 391.50

Merits of Geometric mean :

1. It is rigidly defined
2. It is based on all items
3. It is very suitable for averaging ratios, rates and percentages
4. It is capable of further mathematical treatment.
5. Unlike AM, it is not affected much by the presence of extreme values

Demerits of Geometric mean:

1. It cannot be used when the values are negative or if any of the observations is zero
2. It is difficult to calculate particularly when the items are very large or when there is a frequency distribution.
3. It brings out the property of the ratio of the change and not the absolute difference of change as the case in arithmetic mean.
4. The GM may not be the actual value of the series.

Median

Median is defined as the middle most observation when the observations are arranged in ascending or descending order of magnitude. That means the number of observations preceding median will be equal to the number of observations succeeding it. Median is denoted by M. The median is that value of the variate which divides the group into two equal parts, one part comprising all values greater, and the other, all values less than median

Ungrouped or Raw data :

Arrange the given values in the increasing or decreasing order. If the number of values are odd, median is the middle value. If the number of values are even, median is the mean of middle two values.

By formula

Median (m) = $\frac{N+1}{2}$ th item

Example 10:

When odd number of values are given. Find median for the following data

25, 18, 27, 10, 8, 30, 42, 20, 53

Solution:

Arranging the data in the ascending order 8, 10, 18, 20, 25, 27, 30, 42, 53. The middle value is the 5^{th} item i.e., 25 is the median.

Using this formula

Median (m) = $\frac{N+1}{2}$ th item

Median (m) = $\frac{9+1}{2}$ th item

Median (m) = $\frac{10}{2}$ th item

Median (m) = 5 th item

= 25

Example 11 :

When even number of values are given. Find median for the following data

5, 8, 12, 30, 18, 10, 2, 22

Solution:

Arranging the data in the ascending order 2, 5, 8, 10, 12, 18, 22, 30.

Using the formula

Median (m) = $\frac{N+1}{2}$ th item

Median (m) = $\frac{8+1}{2}$ th item

Median (m) = $\frac{9}{2}$ th item

Median (m) = 4.5 th item

Here median is the mean of the middle two items (ie) mean of (10,12) ie

$$= \frac{value\ \ of\ \ 4^{th}\ \ item + value\ \ of\ \ 5^{th}\ item}{2}$$

$$=(\frac{10+12}{2}) = \frac{22}{2} = 11$$

Example 12:

The following table represents the marks obtained by a batch of 10 students in certain class tests in statistics and Accountancy.

Serial No	1	2	3	4	5	6	7	8	9	10
Marks (Statistics)	53	55	52	32	30	60	47	46	35	28
Marks (Accountancy)	57	45	24	31	25	84	43	80	32	72

Indicate in which subject is the level of knowledge higher?

Solution:

For such question, median is the most suitable measure of central tendency. The mark in the two subjects are first arranged in ascending order as follows:

Serial No	1	2	3	4	5	6	7	8	9	10
Marks in Statistics	28	30	32	35	46	47	52	53	55	60
Marks in Accountancy	24	25	31	32	43	45	57	72	80	84

Median (m) = $\frac{N+1}{2}$ th item

Median (m) = $\frac{10+1}{2}$ th item

Median (m) = $\frac{11}{2}$ th item

Median (m) = 5.5 th item

$\frac{value\ of\ 5^{th}\ item + value\ of\ 6^{th}\ item}{2}$

Median for Statistics = $\frac{46+47}{2}$ = 46.5

Median for Accountancy = $\frac{43+45}{2}$ = 44

Therefore the level of knowledge in Statistics is higher than that in Accountancy.

Grouped Data:

In a grouped distribution, values are associated with frequencies. Grouping can be in the form of a discrete frequency distribution or a continuous frequency distribution. Whatever may be the type of distribution , cumulative frequencies have to be calculated to know the total number of items.

Cumulative frequency: (cf)

Cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the pervious classes, ie adding the frequencies successively, so that the last cumulative frequency gives the total number of items.

Discrete Series:

Step1: Find cumulative frequencies.

Step2: Find $\frac{N+1}{2}$

Step3: See in the cumulative frequencies the value just greater than $\frac{N+1}{2}$

Step4: Then the corresponding value of x is median.

Example 13:

The following data pertaining to the number of members in a family. Find median size of the family.

Number of members x	1	2	3	4	5	6	7	8	9	10	11	12
Frequency F	1	3	5	6	10	13	9	5	3	2	2	1

Solution :

X	**f**	**Cf**
1	1	1
2	3	4
3	5	9
4	6	15
5	10	25
6	13	38
7	9	47
8	5	52
9	3	55
10	2	57
11	2	59
12	1	60
	N=60	

Median (m) = $\frac{N+1}{2}^{th}$ item

Median (m) = $\frac{60+1}{2}$ th item

Median (m) = $\frac{61}{2}$ th item

Median (m) = 30.5 th item

The cumulative frequencies just greater than 30.5 is 38.and the value of x corresponding to 38 is 6.Hence the median size is 6 members per family.

Note:

It is an appropriate method because a fractional value given by mean does not indicate the average number of members in a family.

Continuous Series:

The steps given below are followed for the calculation of median in continuous series.

Step1: Find cumulative frequencies.

Step2: Find $\left(\frac{N}{2}\right)$

Step3: See in the cumulative frequency the value first greater than $(\frac{n}{2})$, Then the corresponding class interval is called the Median class. Then apply the formula

$$median = l + \frac{\frac{N}{2} - cf}{f} \times C$$

Where

l = Lower limit of the median class

cf = cumulative frequency preceding the median

c = width of the median class

f = frequency in the median class.

N =Total frequency.

Note :

If the class intervals are given in inclusive type convert them into exclusive type and call it as true class interval and consider lower limit in this.

Example 14:

The following table gives the frequency distribution of 325 workers of a factory, according to their average monthly income in a certain year.

Income group (in Rs)	Number of workers
Below 100	1
100-150	20
150-200	42
200-250	55
250-300	62
300-350	45
350-400	30
400-450	25
450-500	15
500-550	18
550-600	10
600 and above	2
	325

Calculate median income

Solution :

Income group (in Rs) (Class-interval)	**Number of workers (Frequency)**	**Cumulative frequency c.f**
Below 100	1	1
100-150	20	21
150-200	42	63
200-250	55	118
250-300	62	180
300-350	45	225
350-400	30	255
400-450	25	280
450-500	15	295
500-550	18	313
550-600	10	323
600 and above	2	325
	N= 325	

$\left(\frac{N}{2}\right) = \left(\frac{325}{2}\right)$=162.5

Here

l = 250,

N = 325,

f = 62,

c = 50,

cf = 118

$$\text{median} = l + \frac{\frac{N}{2} - cf}{f} \times c$$

$$\text{median} = 250 + \frac{162.5 - 118}{62} \times 50$$

$$\text{median} = 250 + \frac{44.5}{62} \times 50$$

$$\text{median} = 250 + \frac{162.5 - 118}{62} \times 50$$

$$\text{median} = 250 + \frac{2225}{62}$$

$$\text{median} = 250 + 35.887$$

$$\text{median} = 285.887$$

Example 15:

Calculate median from the following data

Value	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39
Frequency	5	8	10	12	7	6	3	2

Solution

C	f	True class interval	c.f
0-4	5	0.5-4.5	5
5-9	8	4.5-9.5	13
10-14	10	9.5-14.5	23
15-19	12	14.5-19.5	35
20-24	7	19.5-24.5	42
25-29	6	24.5-29.5	48
30-34	3	29.5-34.5	51
35-39	2	34.5-39.5	53
	N=53		

$\left(\frac{N}{2}\right) = \frac{53}{2} = 26.5$

l = 14.5,

N = 53,

f = 12,

c = 5,

cf = 23

$$\text{median} = l + \frac{\frac{N}{2} - cf}{f} \times c$$

$$\text{median} = 14.5 + \frac{26.5 - 23}{12} \times 5$$

$$\text{median} = 14.5 + \frac{3.5}{12} \times 5$$

$$\text{median} = 14.5 + \frac{17.5}{12}$$

$$\text{median} = 14.5 + 1.46$$

$$\text{median} = 15.96$$

Note:

Since the variables are in the inclusive form, classes have to be adjusted. The difference between the upper limit of first class and lower limit of second class is one, in this problem. It is divided by 2, we get 0.5 which have be reduced lower limit of every class limit and have be added with the upper limit of every class.

Example 16:

Compute median for the following data.

Mid-value	5	15	25	35	45	55	65	75
Frequency	7	10	15	17	8	4	6	7

Solution :

The given problem is continuous series frequency distribution. Mid –values of the class limits are given. The difference between two mid – values is 10. Therefore 10/2 or 5 is reduced from each mid value to get the lower limit and 5 is added to get the upper limit of a class.

Mid x	**C.I**	**f**	**c.f**
5	0-10	7	7
15	10-20	10	17
25	20-30	15	32
35	30-40	17	49
45	40-50	8	57
55	50-60	4	61
65	60-70	6	67
75	70-80	7	74
		N=74	

Median = size of $\left(\frac{N}{2}\right) th$ item

$= \frac{74}{2}$ = 37 which lies in the class 30-40

l = 30,

N = 74,

f = 17,

c = 10,

cf = 32

$$median = l + \frac{\frac{N}{2} - cf}{f} \times C$$

$$median = 30 + \frac{37-32}{17} \times 10$$

$$median = 30 + \frac{5}{17} \times 10$$

$$median = 30 + \frac{50}{17}$$

$$median = 30 + 2.94$$

$$median = 32.94$$

Merits of Median :

1. Median is not influenced by extreme values because it is a positional average.
2. Median can be calculated in case of distribution with open-end intervals.
3. Median can be located even if the data are incomplete.
4. Median can be located even for qualitative factors such as ability, honesty etc.

Demerits of Median :

1. A slight change in the series may bring drastic change in median value.
2. In case of even number of items or continuous series, median is an estimated value other than any value in the series.
3. It is not suitable for further mathematical treatment except its use in mean deviation.
4. It is not taken into account all the observations

Mode

The mode refers to that value in a distribution, which occur most frequently. It is an actual value, which has the highest concentration of items in and around it. According to Croxton and Cowden " The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded at the most typical of a series of values".

It shows the centre of concentration of the frequency in around a given value. Therefore, where the purpose is to know the point of the highest concentration it is preferred. It is, thus, a positional measure.

Its importance is very great in marketing studies where a manager is interested in knowing about the size, which has the highest concentration of items. For example, in placing an order for shoes or ready-made garments the modal size helps because this sizes and other sizes around in common demand.

Computation of the mode:

Ungrouped or Raw Data:

For ungrouped data or a series of individual observations, mode is often found by mere inspection.

Example 17:

2 , 7, 10, 15, 10, 17, 8, 10, 2

$\therefore$ Mode = M_0 = 10

In some cases the mode may be absent while in some cases there may be more than one mode.

Example 18:

1) 12, 10, 15, 24, 30 (no mode)

2) 7, 10, 15, 12, 7, 14, 24, 10, 7, 20, 10

the modes are 7 and 10

Grouped Data:

For Discrete distribution, see the highest frequency and corresponding value of X is mode.

Continuous distribution:

See the highest frequency then the corresponding value of class interval is called the modal class. Then apply the formula.

Mode = $M_{0=}\ l + \frac{\Delta_1}{\Delta_{1+}\Delta_2} \times C$

l = Lower limit of the model class

$\Delta_1 = f_1 - f_0$

$\Delta_2 = f_1 - f_2$

f_1 = frequency of the modal class

f_0 = frequency of the class preceding the modal class

f_2 = frequency of the class succeeding the modal class

Example 19:

Calculate mode for the following:

C.I	F
0-50	5
50-100	14
100-150	40
150-200	91
200-250	150
250-300	87
300-350	60
350-400	38
400 and above	15

Solution:

The highest frequency is 150 and corresponding class interval is 200 – 250, which is the modal class.

l = Lower limit of the model class 200

f_1 = frequency of the modal class 150

f_0 = frequency of the class preceding the modal class 91

f_2 = frequency of the class succeeding the modal class 87

$\Delta_1 = f_1 - f_0$ 150-91 = 59

$\Delta_2 = f_1 - f_2$ 150-87 =63

C=50

Mode = $M_{0} = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$

$= 200 + \frac{59}{59+63} \times 50$

$= 200 + \frac{2950}{122}$

=200+ 24.18

= 224.18

Exercises

1. State the merits of arithmetic mean as a measure of central tendency
2. What are the main characteristics of a good mean?
3. Write the advantages and disadvantages of Harmonic mean
4. Write the advantages and disadvantages of Geometric mean
5. What is meant by median? explain
6. What is meant by mode? Explain

Marks	10	20	30	40	50	60
No of students	5	6	10	8	5	6

7. Calculate arithmetic mean from the following

(Mean=35)

8. Calculate arithmetic mean from the following

Marks	0-10	10-20	20-30	30-40	40-50
Frequency	5	8	10	8	9

(Mean=27)

9. Calculate median from the following

50,100,300,500,600,1000,1150 ***(Median =500)***

10. Calculate median from the following

50,100,300,500,600,700,1000,1150 ***(Median =550)***

11. Calculate median from the following

Marks	10	20	30	40	50
No of students	2	5	10	4	3

(Median =30)

12. Calculate median from the following

Class interval	10-15	15-20	20-25	25-30	30-35
Frequency	10	14	20	13	8

(Median =22.125)

13. Calculate mode of the following

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No of students	3	6	10	15	8	8

(Mode=34.17)

14. Calculate mean and median

Mid point	:	400	600	800	1000	1200	1400
Frequency	:	25	55	30	20	14	6

$(\bar{x} = 748, m = 681.81)$

15. Find Harmonic mean

X	:	10	12	14	16	18	20
Y	:	5	18	20	10	6	1

(HM=13.42)

16. Find out the Harmonic mean

X	:	1	2	3	4	5
Y	:	2	4	3	2	1

(HM=2.10)

17. Calculate the mean, median and mode for the following frequency distribution

Class internal	:	0-4	5-9	10-14	15-19	20-24	25-29	30-34
Frequency	:	4	7	15	18	12	10	3

(Mean =17, Median=16.86, Mode=16.7)

18. Calculate the mean, median and mode for the following frequency distribution

Mid point	5	10	15	20	25	30
Frequency	5	15	25	30	15	10

(Mean =18.5, Median=19.17, Mode=20)

Introduction

The measure of central tendency serves to locate the center of the distribution, but they do not reveal how the items are spread out on either side of the center. This characteristic of a frequency distribution is commonly referred to as dispersion. In a series all the items are not equal. There is difference or variation among the values. The degree of variation is evaluated by various measures of dispersion. Small dispersion indicates high uniformity of the items, while large dispersion indicates less uniformity. For example consider the following marks of two students.

Student I	Student II
68	85
75	90
65	80
67	25
70	65

Both have got a total of 345 and an average of 69 each. The fact is that the second student has failed in one paper. When the averages alone are considered, the two students are equal. But first student has less variation than second student. Less variation is a desirable characteristic.

Characteristics of a good measure of dispersion

An ideal measure of dispersion is expected to possess the following properties

1. It should be rigidly defined
2. It should be based on all the items.
3. It should not be unduly affected by extreme items.
4. It should lend itself for algebraic manipulation.
5. It should be simple to understand and easy to calculate

Absolute and Relative Measures:

There are two kinds of measures of dispersion, namely

1. Absolute measure of dispersion
2. Relative measure of dispersion.

Absolute measure of dispersion indicates the amount of variation in a set of values in terms of units of observations. For example, when rainfalls on different days are available in mm, any absolute measure of dispersion gives the variation in rainfall in mm. On the other hand relative measures of dispersion are free from the units of measurements of the observations They are pure numbers. They are used to compare the variation in two or more sets, which are having different units of measurements of observations.

The various absolute and relative measures of dispersion are listed below.

Absolute measure	**Relative measure**
1. Range	1. Co-efficient of Range
2. Quartile deviation	2. Co-efficient of Quartile deviation
3. Mean deviation	3. Co-efficient of Mean deviation
4. Standard deviation	4. Co-efficient of variation

Range and coefficient of Range

Range:

This is the simplest possible measure of dispersion and is defined as the difference between the largest and smallest values of the variable.

In symbols, Range = L – S.

Where L = Largest value.

S = Smallest value.

In individual observations and discrete series, L and S are easily identified. In continuous series, the following two methods are followed.

Method 1:

L = Upper boundary of the highest class

S = Lower boundary of the lowest class.

Method 2:

L = Mid value of the highest class.

S = Mid value of the lowest class.

Co-efficient of Range:

Co-efficient of Range = $\frac{L-S}{L+S}$

Example 1:

Find the value of range and its co-efficient for the following data.

7, 9, 6, 8, 11, 10

Solution:

L = 11, S = 4.

Range = L – S = 11- 4 = 7

Co-efficient of Range = $\frac{L-S}{L+S}$

Co-efficient of Range = $\frac{11-4}{11+4}$

Co-efficient of Range = $\frac{7}{15}$

Co-efficient of Range = 0.4667

Example 2:

Calculate range and its co efficient from the following distribution.

Size:	60-63	63-66	66-69	69-72	72-75
Number:	5	18	42	27	8

Solution:

L = Upper boundary of the highest class. = 75

S = Lower boundary of the lowest class. = 60

Range = L – S = 75 – 60 = 15

Co-efficient of Range = $\frac{L-S}{L+S}$

Co-efficient of Range = $\frac{75-60}{75+60}$

Co-efficient of Range = $\frac{15}{135}$

Co-efficient of Range = 0.1111

Merits and Demerits of Range:

Merits:

1. It is simple to understand.
2. It is easy to calculate.
3. In certain types of problems like quality control, weather forecasts, share price analysis, etc., range is most widely used.

Demerits:

1. It is very much affected by the extreme items.
2. It is based on only two extreme observations.
3. It cannot be calculated from open-end class intervals.
4. It is not suitable for mathematical treatment.
5. It is a very rarely used measure.

Quartile Deviation and Co efficient of Quartile Deviation

Quartile Deviation (Q.D):

Definition:

Quartile Deviation is half of the difference between the first and third quartiles.Hence, it is called Semi Inter Quartile Range.

In Symbols, Q .D = $\frac{Q_3-Q_1}{2}$. Among the quartiles Q1, Q2 and Q3, the range Q3 – Q1 is called inter quartile range and $\frac{Q_3-Q_1}{2}$. , semi- inter quartile range.

Co-efficient of Quartile Deviation :

Co-efficient of Quartile Deviation = $\frac{Q_3-Q_1}{Q_3+Q_1}$.

Example 3:

Find the Quartile Deviation for the following data: 391, 384, 591, 407, 672, 522, 777, 733, 1490, and 2488

Solution:

Arrange the given values in ascending order.

384, 391, 407, 522, 591, 672, 733, 777, 1490, 2488.

Position of Q_1 is $\frac{N+1}{4} = \frac{10+1}{4} = \frac{11}{4} = 2.75^{\text{th item}}$

Q1 = 2nd value + 0.75 (3rd value – 2nd value)

= 391 + 0.75 (407 – 391)

= 391 + 0.75 × 16

= 391 + 12

= 403

Position of Q3 is $3\left(\frac{N+1}{4}\right) = 3 \times 2.75 = 8.25$

Q3 = 8th value + 0.25 (9th value – 8th value)

= 777 + 0.25 (1490 – 777)

= 777 + 0.25 (713)

= 777 + 178.25 = 955.25

$Q.D = \frac{Q_3 - Q_1}{2}.$

$Q.D = \frac{955.25 - 403}{2}.$

$Q.D = \frac{552.25}{2}.$

Q.D = 276.125

Example 4 :

Weekly wages of labours are given below. Calculate Q.D and Coefficient of Q.D.

Weekly Wage (Rs.)	:	100	200	400	500	600
No. of Weeks	:	5	8	21	12	6

Solution:

Weekly Wage (Rs.)	No. of Weeks	Cum. No. of Weeks
100	5	5
200	8	13
400	21	34
500	12	46
600	6	52
Total	**N = 52**	

Position of Q_1 is $\frac{N+1}{4} = \frac{52+1}{4} = \frac{53}{4} = 13.25^{\text{th item}}$

Q_1 = 13^{th} value + 0.25 (14^{th} Value – 13^{th} value)

= 13^{th} value + 0.25 (400 – 200)

= 200 + 0.25 (400 – 200)

= 200 + 0.25 (200)

= 200 + 50 = 250

Position of Q3 is $3\left(\frac{N+1}{4}\right)$ = 3 × 13.25 = 39.75

Q3 = 39th value + 0.75 (40th value – 39th value)

= 500 + 0.75 (500 – 500)

= 500 + 0.75 × 0

= 500

$Q.D = \frac{Q_3 - Q_1}{2}.$

$Q.D = \frac{500 - 250}{2}.$ $Q.D = \frac{250}{2}.$ $Q.D = 125$

Co-efficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}.$

Co-efficient of Q.D = $\frac{500 - 250}{500 + 250}.$; **Co-efficient of Q.D =** $\frac{250}{750}.$

Co-efficient of Quartile Deviation = 0.3333

Example 5:

For the data given below, give the quartile deviation and coefficient of quartile deviation.

X :	351 – 500	501 – 650	651 – 800	801–950	951–1100
f :	48	189	88	47	28

Solution :

X	F	True class Intervals	Cumulative frequency
351-500	48	350.5-500.5	48
501-650 (Q_1)	189	500.5-650.5	237
651-800 (Q_3)	88	650.5-800.5	325
801-950	47	800.5-950.5	372
951-1100	28	950.5-1100.5	400
Total	**N = 400**		

$$Q_1 = l \pm \frac{\frac{N}{4} - cf}{f} \times C$$

$\frac{N}{4} = \frac{400}{4} =$ 100 which lies in 500.5-650.5

l=500.5

cf= 48

f=189

c=150

$$Q_1 = 500.5 + \frac{100 - 48}{189} \times 150$$

$$Q_1 = 500.5 + \frac{52}{189} \times 150$$

$$Q_1 = 500.5 + \frac{52 \times 150}{189}$$

$$Q_1 = 500.5 + \frac{7800}{189}$$

$$Q_1 = 500.5 + 41.27$$

$$Q_1 = 541.77$$

$$Q_3 = l \pm \frac{3\left(\frac{N}{4}\right) - cf}{f} \times C$$

$3\left(\frac{N}{4}\right) = 3 \times 100 = 300$ which lies in 650.5-800.5

Q_3 Class is 650.5 – 800.5

l = 650.5,

cf = 237,

f = 88,

C = 150

$$Q_3 = 650.5 + \frac{300 - 237}{88} \times 150$$

$$Q_3 = 650.5 + \frac{63}{88} \times 150$$

$$Q_3 = 650.5 + \frac{63 \times 150}{88}$$

$$Q_3 = 650.5 + 107.39$$

$Q_3 = 751.89$

Therefore

$Q.D = \frac{Q_3 - Q_1}{2}$.

$$Q.D = \frac{751.89 - 541.77}{2}$$

$$Q.D = \frac{216.12}{2}$$

$$Q.D = 108.06$$

Co-efficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$.

$$\frac{751.89 - 541.77}{751.89 + 541.77}$$

Co-efficient of Quartile Deviation = $\frac{216.12}{1299.66}$.

Co-efficient of Quartile Deviation = 0.1663

Merits and Demerits of Quartile Deviation

Merits :

1. It is Simple to understand and easy to calculate
2. It is not affected by extreme values.
3. It can be calculated for data with open end classes also.

Demerits:

1. It is not based on all the items. It is based on two positional values Q_1 and Q_3 and ignores the extreme 50% of the items
2. It is not amenable to further mathematical treatment.
3. It is affected by sampling fluctuations.

Mean Deviation and Coefficient of Mean Deviation

Mean Deviation:

The range and quartile deviation are not based on all observations. They are positional measures of dispersion. They do not show any scatter of the observations from an average. The mean deviation is measure of dispersion based on all items in a distribution.

Definition:

Mean deviation is the arithmetic mean of the deviations of a series computed from any measure of central tendency; i.e., the mean, median or mode, all the deviations are taken as positive i.e., signs are ignored. According to Clark and Schekade,

"Average deviation is the average amount scatter of the items in a distribution from either the mean or the median, ignoring the signs of the deviations".

We usually compute mean deviation about any one of the three averages mean, median or mode. Sometimes mode may be ill defined and as such mean deviation is computed from mean and median. Median is preferred as a choice between mean and median. But in general practice and due to wide applications of mean, the mean deviation is generally computed from mean. M.D can be used to denote mean deviation.

Coefficient of mean deviation:

Mean deviation calculated by any measure of central tendency is an absolute measure. For the purpose of comparing variation among different series, a relative mean deviation is required. The relative mean deviation is obtained by dividing the mean deviation by the average used for calculating mean deviation.

$$c\text{o efficient of mean deviation} = \frac{\text{mean deviation}}{\text{mean or median or mode}}$$

If the result is desired in percentage, the coefficient of mean deviation

$$\text{co efficient of mean deviation} \frac{\text{mean deviation}}{\text{mean or median or mode}} \times 100$$

Computation of mean deviation – Individual Series:

1. Calculate the average mean, median or mode of the series.
2. Take the deviations of items from average ignoring signs and denote these deviations by |D|.

3. Compute the total of these deviations, i.e., $\sum |D|$

4. Divide this total obtained by the number of items.

Symbolically: $M.D = \frac{\sum |D|}{n}$

Example 6:

Calculate mean deviation from mean and median for the following data:

100,150,200,250,360,490,500,600,671 also calculate coefficients of M.D.

Solution:

MEAN

$$\bar{X} = \frac{\sum X}{N} \quad ; \quad \bar{X} = \frac{3321}{9} = 369$$

MEDIAN

Now arrange the data in ascending order

100, 150, 200, 250, 360, 490, 500, 600, 671

$Median = \left(\frac{N+1}{2}\right)^{th}$ item ; $Median = \left(\frac{9+1}{2}\right)^{th}$ item $Median = \left(\frac{10}{2}\right)^{th}$ item ;

$Median = (5)^{th}$ item $Median = 360$

X	$\|D\| = \|X - \bar{X}\|$	$\|D\| = \|X - median\|$
100	269	260
150	219	210
200	169	160
250	119	110
360	9	0
490	121	130
500	131	140

600	231	240
671	302	311
3321	$\sum\|\boldsymbol{D}\|$**=1570**	$\sum\|\boldsymbol{D}\|$**=1561**

$$\text{M. D from mean} = \frac{\sum|\text{D}|}{\text{n}}$$

$$\text{M. D from mean} = \frac{1570}{9} = 174.44$$

$$\text{co efficient of mean deviation} = \frac{\text{mean deviation}}{\text{mean}}$$

$$\text{co efficient of mean deviation} = \frac{174.44}{369}$$

$$\text{co efficient of mean deviation} = 0.47$$

$$\text{M. D from median} = \frac{\sum|\text{D}|}{\text{n}}$$

$$M.D = \frac{1561}{9} = 173.44$$

$$co\ efficient\ of\ mean\ deviation = \frac{mean\ deviation}{median}$$

$$co\ efficient\ of\ mean\ deviation = \frac{173.44}{360} = 0.48$$

Mean Deviation – Discrete series:

Steps: 1. Find out an average (mean, median or mode).

2. Find out the deviation of the variable values from the average, ignoring signs and denote them by |D|

3. Multiply thedeviation of each value by its respective frequency and find out the total $\sum f \mid D|$

4. Divide Σf | D| by the total frequencies N

Symbolically, M.D= $\frac{\boldsymbol{\Sigma f \mid D|}}{\boldsymbol{N}}$

Example 7:

Compute Mean deviation from mean and median from the following data:

Height in cms	158	159	160	161	162	163	164	165	166
No. of persons	15	20	32	35	33	22	20	10	8

Also compute coefficient of mean deviation

Solution :

Height X	**No. of persons F**	**d = x-A A = 162**	**Fd**	**|D| = |X-mean|**	**f|D|**
158	15	-4	-60	3.51	52.65
159	20	-3	-60	2.51	50.20
160	32	-2	-64	1.51	48.32
161	35	-1	-35	0.51	17.85
162	33	0	0	0.49	16.17
163	22	1	22	1.49	32.78
164	20	2	40	2.49	49.80
165	10	3	30	3.49	34.90
166	8	4	32	4.49	35.92
	N=195		**∑fd=-95**		**∑f|D|=338.59**

$\bar{X} = A \pm \frac{\sum fd}{N}$

$\bar{X} = 162 - \frac{95}{195}$

$\bar{X} = 162 - 0.49$

$\bar{X} = 161.51$

$$\text{M. D from mean} = \frac{\sum|\text{D}|}{\text{n}}$$

$$\text{M. D from mean} = \frac{338.59}{195}$$

$$\text{M. D from mean} = 1.74$$

$$\text{co efficient of mean deviation} = \frac{\text{mean deviation}}{\text{mean}}$$

$$\text{co efficient of mean deviation} = \frac{1.74}{161.51}$$

$$\text{co efficient of mean deviation} = 0.0108$$

Height x	No. of persons f	c.f.	|D| = |X – Median|	f |D|
158	15	15	3	45
159	20	35	2	40
160	32	67	1	32
161	35	102	0	0
162	33	135	1	33
163	22	157	2	44
164	20	177	3	60
165	10	187	4	40
166	8	195	5	40
	N=195			**∑ f |D|=334**

Median (m) = $\frac{N+1}{2}$ th item

Median (m) = $\frac{195+1}{2}$ th item

Median (m) = 98 th item

Median (m) = 161

$$\text{M. D from median} = \frac{\sum|D|}{n}$$

$$\text{M. D from median} = \frac{334}{195}$$

$$\text{M. D from median} = 1.71$$

$$\text{co efficient of mean deviation} = \frac{\text{mean deviation}}{\text{median}}$$

$$\text{co efficient of mean deviation} = \frac{1.71}{161}$$

$$\text{co efficient of mean deviation} = 0.0106$$

Mean deviation-Continuous series:

The method of calculating mean deviation in a continuous series same as the discrete series. In continuous series we have to find out the mid points of the various classes and take deviation of these points from the average selected. Thus

M.D= $\frac{\boldsymbol{\Sigma f \mid D|}}{\boldsymbol{N}}$

Where

D = M – average

M = Mid point

Example 8:

Find out the mean deviation from mean and median from the following series.

Age in years	No. of persons
0-10	20
10-20	25
20-30	32
30-40	40
40-50	42
50-60	35
60-70	10
70-80	8

Also compute co-efficient of mean deviation.

Solution:

X	M	F	d = $\frac{M-A}{C}$ (A=35, C=10)	Fd	$\mid D \mid = \mid m - \bar{x} \mid$	f $\mid D \mid$
0-10	5	20	-3	-60	31.5	630.0
10-20	15	25	-2	-50	21.5	537.5
20-30	25	32	-1	-32	11.5	368.0
30-40	**35**	40	0	0	1.5	60.0
40-50	45	42	1	42	8.5	357.0

50-60	55	35	2	70	18.5	647.5
60-70	65	10	3	30	28.5	285.0
70-80	75	8	4	32	**38.5**	308.0
		N=212		**∑fd=32**		**∑f|D|=3192.5**

$\bar{X} = A \pm \frac{\sum fd}{N} \times c$

$\bar{X} = 35 + \frac{32}{212} \times 10$

$\bar{X} = 35 + \frac{320}{212}$

$\bar{X} = 35 + 1.51$

$\bar{X} = 36.51$

$\text{M. D from mean} = \frac{\sum|\text{D}|}{\text{n}}$

$\text{M. D from mean} = \frac{3192.5}{212}$

$\text{M. D from mean} = 15.06$

$\text{co efficient of mean deviation} = \frac{\text{mean deviation}}{\text{mean}}$

$\text{co efficient of mean deviation} = \frac{15.06}{36.51}$

$\text{co efficient of mean deviation} = 0.41$

Calculation of median and M.D. from median

X	M	F	c.f	|D| = |m-Md|	f | D |
0-10	5	20	20	32.25	645.00
10-20	15	25	45	22.25	556.25
20-30	25	32	77	12.25	392.00
30-40	35	40	117	2.25	90.00
40-50	45	42	159	7.75	325.50
50-60	55	35	194	17.75	621.25
60-70	65	10	204	27.75	277.50
70-80	75	8	212	37.75	302.00
		N=212		**Total**	**3209.50**

$$median = l + \frac{\frac{N}{2} - cf}{f} \times C$$

$\frac{N}{2} = \frac{212}{2} =$106 which lies in 30-40

l=30

cf=77

f=40

c=10

$$median = 30 + \frac{106 - 77}{40} \times 10$$

$$median = 30 + \frac{29}{40} \times 10$$

$$median = 30 + \frac{290}{40}$$

$$median = 30 + 7.25.$$

$$median = 37.25$$

$$\text{M. D from median} = \frac{\sum|\text{D}|}{\text{n}}$$

$$\text{M. D from median} = \frac{3209.5}{212} = 15.14$$

$$\text{co efficient of mean deviation} = \frac{\text{mean deviation}}{\text{median}}$$

$$\text{co efficient of mean deviation} = \frac{15.14}{37.25}$$

$$\text{co efficient of mean deviation} = 0.41$$

Merits and Demerits of M.D :

Merits:

1. It is simple to understand and easy to compute.
2. It is rigidly defined.
3. It is based on all items of the series.
4. It is not much affected by the fluctuations of sampling.
5. It is less affected by the extreme items.
6. It is flexible, because it can be calculated from any average.
7. It is better measure of comparison.

Demerits:

1. It is not a very accurate measure of dispersion.
2. It is not suitable for further mathematical calculation.
3. It is rarely used. It is not as popular as standard deviation.

Algebraic positive and negative signs are ignored. It is mathematically unsound and illogical.

Standard Deviation and Coefficient of variation

Standard Deviation:

Karl Pearson introduced the concept of standard deviation in 1893. It is the most important measure of dispersion and is widely used in many statistical formulae. Standard deviation is also called Root-Mean Square Deviation. The reason is that it is the square–root of the mean of the squared deviation from the arithmetic mean. It provides accurate result. Square of standard deviation is called Variance.

Definition:

It is defined as the positive square-root of the arithmetic mean of the Square of the deviations of the given observation from their arithmetic mean. The standard deviation is denoted by the Greek letter **σ** (sigma)

Calculation of Standard deviation-Individual Series:

There are two methods of calculating Standard deviation in an individual series.

a) Deviations taken from Actual mean

b) Deviation taken from Assumed mean

a) Deviation taken from Actual mean:

This method is adopted when the mean is a whole number.

Steps:

1. Find out the actual mean of the series (x)
2. Find out the deviation of each value from the mean (x = X – $\bar{X}$)
3. Square the deviations and take the total of squared deviations $\sum x^2$
4. Divide the total ($\sum X^2$) by the number of observation $\left(\frac{\sum X^2}{N}\right)$

The square root of $\left(\frac{\sum X^2}{N}\right)$ is standard deviation.

Thus **σ =** $\sqrt{\frac{\sum X^2}{N}}$ **or** $\sqrt{\frac{\sum(X-\bar{X})}{\boldsymbol{N}}}$

a) Deviations taken from assumed mean:

This method is adopted when the arithmetic mean is fractional value. Taking deviations from fractional value would be a very difficult and tedious task. To save time and labour, we apply short–cut method; deviations are taken from an assumed mean. The formula is:

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

Where d-stands for the deviation from assumed mean = (X-A)

Steps:

1. Assume any one of the item in the series as an average (A)
2. Find out the deviations from the assumed mean; i.e., X-A denoted by d and also the total of the deviations $\sum d$
3. Square the deviations; i.e., d^2 and add up the squares of deviations, i.e, $\sum d^2$
4. Then substitute the values in the following formula:

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

Example 9:

Calculate the standard deviation from the following data.

14, 22, 9, 15, 20, 17, 12, 11

Solution:

Deviations from actual mean.

Values (X)	**X - X**	**$(X-X)^2$**
14	-1	1
22	7	49
9	-6	36
15	0	0
20	5	25
17	2	4
12	-3	9
11	-4	16
N=120		**$\sum(X-X)^2=140$**

$$\sigma = \sqrt{\frac{\sum(X-\bar{X})}{N}}$$

$$\sigma = \sqrt{\frac{\sum(X-\bar{X})}{N}}$$

$$\sigma = \sqrt{\frac{140}{8}}$$

$$\sigma = \sqrt{17.5}$$

$$\sigma = 4.18$$

Example 10:

The table below gives the marks obtained by 10 students in statistics. Calculate standard deviation.

Student Nos :	1	2	3	4	5	6	7	8	9	10
Marks :	43	48	65	57	31	60	37	48	78	59

Solution: (Deviations from assumed mean)

Nos.	**Marks (x)**	**d=X-A (A=57)**	d^2
1	43	-14	196
2	48	-9	81
3	65	8	64
4	57	0	0
5	31	-26	676
6	60	3	9
7	37	-20	400
8	48	-9	81
9	78	21	441
10	59	2	4
N = 10		$\sum d = 44$	$\sum d^2 = 1952$

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

$$\boldsymbol{\sigma} = \sqrt{\frac{1952}{10} - \left(\frac{44}{10}\right)^2}$$

$$\boldsymbol{\sigma} = \sqrt{195.2 - 19.36}$$

$$\sigma = \sqrt{175.84}$$

$\sigma = 13.26$

Calculation of standard deviation:

Discrete Series:

There are three methods for calculating standard deviation in discrete series:

(a) Actual mean methods

(b) Assumed mean method

(c) Step-deviation method.

(a) Actual mean method:

Steps:

1. Calculate the mean of the series.

2. Find deviations for various items from the means i.e., $x - \bar{X} = d$.

3. Square the deviations ($=d^2$) and multiply by the respective frequencies(f) we get fd^2

4. Total to product ($\sum fd^2$) Then apply the formula:

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

If the actual mean in fractions, the calculation takes lot of time and labour; and as such this method is rarely used in practice.

(b) Assumed mean method:

Here deviations are taken not from an actual mean but from an assumed mean. Also this method is used, if the given variable values are not in equal intervals.

Steps:

1. Assume any one of the items in the series as an assumed mean and denoted by A.

2. Find out the deviations from assumed mean, i.e., X-A and denote it by d.

3. Multiply these deviations by the respective frequencies and get the ∑fd.

4. Square the deviations (d^2).

5. Multiply the squared deviations (d^2) by the respective frequencies (f) and get $\sum fd^2$.

6. Substitute the values in the following formula:

$$\boldsymbol{\sigma} = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

where d = X – A, N = ∑f.

Example 11:

Calculate Standard deviation from the following data.

X :	20	22	25	31	35	40	42	45
f :	5	12	15	20	25	14	10	6

Solution :

Deviations from assumed mean

x	F	d = x –A (A = 31)	d^2	fd	fd^2
20	5	-11	121	-55	605
22	12	-9	81	-108	972
25	15	-6	36	-90	540
31	20	0	0	0	0
35	25	4	16	100	400
40	14	9	81	126	1134
42	10	11	121	110	1210
45	6	14	196	84	1176
	N = 107			**∑fd= 167**	**$\sum fd^2$ = 6037**

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

$$\sigma = \sqrt{\frac{6037}{107} - \left(\frac{167}{107}\right)^2}$$

$$\sigma = \sqrt{56.42 - 2.44}$$

$$\sigma = \sqrt{53.98}$$

σ =7.35

(c) Step-deviation method:

If the variable values are in equal intervals, then we adopt this method.

Steps:

1. Assume the center value of the series as assumed mean A.
2. Find $d' = \frac{X-A}{C}$, where C is the interval between each value.
3. Multiply these deviations d' by the respective frequencies and get $\sum fd^t$.
4. Square the deviations and get d'^2.
5. Multiply the squared deviation (d'^2) by the respective frequencies (f) and obtain the total $\sum fd'^2$.
6. Substitute the values in the following formula to get the standard deviation.

$$\sigma = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} \times C$$

Example 12:

Compute Standard deviation from the following data

Marks :	10	20	30	40	50	60
No. of students :	8	12	20	10	7	3

Solution:

Marks x	F	$d' = \frac{X - A}{C}$	fd'	fd'^2
10	8	-2	-16	32
20	12	-1	-12	12
30	20	0	0	0
40	10	1	10	10
50	7	2	14	28
60	3	3	9	27
	N = 60		**$\sum fd' = 5$**	**$\sum fd'^2 = 109$**

$$\sigma = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} \times C$$

$$\sigma = \sqrt{\frac{109}{60} - \left(\frac{5}{60}\right)^2} \times 10$$

$$\sigma = \sqrt{1.817 - 0.0069} \times 10$$

$$\sigma = \sqrt{1.8101} \times 10$$

$$\sigma = 1.345 \times 10$$

$$\sigma = 13.45$$

Calculation of Standard Deviation –Continuous series:

In the continuous series the method of calculating standard deviation is almost the same as in a discrete series. But in a continuous series, mid-values of the class intervals are to be found out. The step- deviation method is widely used. The formula is,

$$\sigma = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} \times C$$

Where

$d' = \frac{X-A}{C}$,

C is the interval between each value.

Steps:

1. Find out the mid-value of each class.
2. Assume the center value as an assumed mean and denote it by A.
3. Find out $d' = \frac{X-A}{C}$,
4. Multiply the deviations d' by the respective frequencies and get ∑fd'.
5. Square the deviations and get d'^2.
6. Multiply the squared deviations (d'^2) by the respective frequencies and get $\sum fd'^2$.
7. Substituting the values in the following formula to get the standard deviation

$$\sigma = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} \times C$$

Example 13:

The daily temperature recorded in a city in Russia in a year is given below.

Temperature C^0	No. of days
– 40 to – 30	10
– 30 to – 20	18
– 20 to – 10	30
– 10 to 0	42
0 to 10	65
10 to 20	180
20 to 30	20
	365

Calculate Standard Deviation.

Solution :

Temperature	Mid value (m)	No. of days f	$d' = \frac{X - A}{C}$	fd'	fd'^2
- 40 to - 30	-35	10	-3	-30	90
-30 to - 20	-25	18	-2	-36	72
-20 to -10	-15	30	-1	-30	30
-10 to -0	-5	42	0	0	0
0 to 10	5	65	1	65	65
10 to 20	15	180	2	360	720
20 to 30	25	20	3	60	180
		N = 365		$\sum fd' = 389$	$\sum fd'^2 = 1157$

$$\sigma = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} \times C$$

$$\sigma = \sqrt{\frac{1157}{365} - \left(\frac{389}{365}\right)^2} \times 10$$

$$\sigma = \sqrt{3.1699 - 1.1358} \times 10$$

$$\sigma = \sqrt{2.0341} \times 10$$

$$\sigma = 1.4262 \times 10$$

$$\sigma = 14.262$$

Merits and Demerits of Standard Deviation:

Merits:

1. It is rigidly defined and its value is always definite and based on all the observations and the actual signs of deviations are used.
2. As it is based on arithmetic mean, it has all the merits of arithmetic mean.
3. It is the most important and widely used measure of dispersion.
4. It is possible for further algebraic treatment.

5. It is less affected by the fluctuations of sampling and hence stable.
6. It is the basis for measuring the coefficient of correlation and sampling.

Demerits:

1. It is not easy to understand and it is difficult to calculate.
2. It gives more weight to extreme values because the values are squared up.
3. As it is an absolute measure of variability, it cannot be used for the purpose of comparison.

Coefficient of Variation:

The Standard deviation is an absolute measure of dispersion. It is expressed in terms of units in which the original figures are collected and stated. The standard deviation of heights of students cannot be compared with the standard deviation of weights of students, as both are expressed in different units, i.e. heights in centimetre and weights in kilograms. Therefore the standard deviation must be converted into a relative measure of dispersion for the purpose of comparison. The relative measure is known as the coefficient of variation. The coefficient of variation is obtained by dividing the standard deviation by the mean and multiplies it by 100.

Symbolically,

Coefficient of variation (C.V) $= \frac{\sigma}{X} \times 100$

If we want to compare the variability of two or more series, we can use C.V. The series or groups of data for which the C.V. is greater indicate that the group is more variable, less stable, less uniform, less consistent or less homogeneous. If the C.V. is less, it indicates that the group is less variable, more stable, more uniform, more consistent or more homogeneous.

Exercises

1. What is quartile deviation
2. Write short notes about mean deviation.

3. Calculate quartile deviation and its co –efficient from the following data

Marks	10	20	30	40	50	60	70	80
No of students	6	8	10	20	12	9	7	5

(Q.D=15, Coefficient of Q.D=0.333)

4. From the following data, calculate quartile deviation

Wages	35-40	40-45	45-50	50-55	55-60
Workers	2	14	24	30	15

(Q.D=3.934)

5. Calculate the mean and standard deviation from the following data

Mid point	:	7.5	12.5	17.5	22.5	27.5	32.5	37.5	42.5
Frequency	:	5	6	15	10	5	4	3	2

(Mean =21.3 and S.D=8.975)

6. Calculate the standard deviation from the following data

   14, 22, 9, 15, 20,17,12,11 ***(S.D=4.18)***

7. Calculate the mean deviation and its co-efficient from the following data

Marks	22-25	25-28	28-31	31-34	34-37	37-40
No of students	4	8	12	3	2	1

(Median=28.75, M.D=2.75, Coefficient of M.D=0.196)

8. Calculate mean deviation from the median from the following data

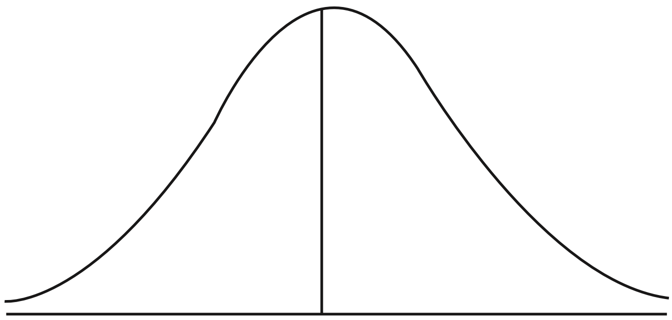
Marks	0-10	10-20	20-30	30-40	40-50
No of students	1	3	8	2	2

(Median=25, M.D=6.875, Coefficient of M.D=0.275)

Introduction

Skewness means 'lack of symmetry'. We study skewness to have an idea about the shape of the curve which we can draw with the help of the given data.If in a distribution mean = median = mode, then that distribution is known as symmetrical distribution. If in a distribution mean ≠ median ≠ mode , then it is not a symmetrical distribution and it is called a skewed distribution and such a distribution could either be positively skewed or negatively skewed.

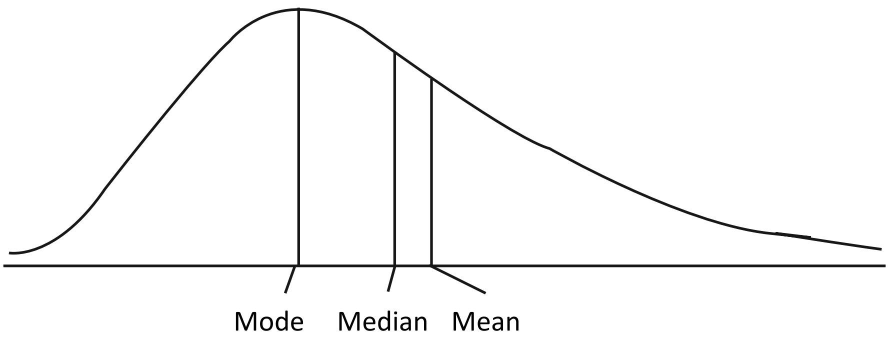
a) Symmetrical distribution:



Mean = Median = Mode

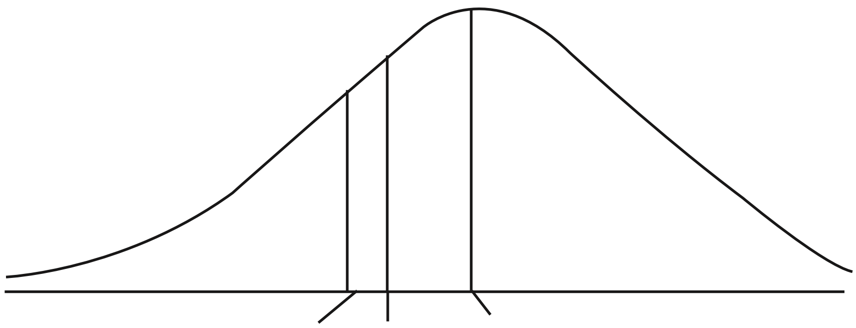
It is clear from the above diagram that in a symmetrical distribution the values of mean, median and mode coincide. The spread of the frequencies is the same on both sides of the center point of the curve.

b) Positively skewed distribution:



It is clear from the above diagram, in a positively skewed distribution, the value of the mean is maximum and that of the mode is least, the median lies in between the two. In the positively skewed distribution the frequencies are spread out over a greater range of values on the right hand side than they are on the left hand side.

c) Negatively skewed distribution:



It is clear from the above diagram, in a negatively skewed distribution, the value of the mode is maximum and that of the mean is least. The median lies in between the two. In the negatively skewed distribution the frequencies are spread out over a greater range of values on the left hand side than they are on the right hand side.

Measures of skewness

The important measures of skewness are

(i) Karl –Pearason's coefficient of skewness

(ii) Bowley's coefficient of skewness

(iii) Measure of skewness based on moments

Karl – Pearson' s Coefficient of skewness:

According to Karl – Pearson, the absolute measure of skewness = ***mean – mode***. This measure is not suitable for making valid comparison of the skewness in two or more distributions because the unit of measurement may be different in different series. To avoid this difficulty use relative measure of skewness called Karl–Pearson' s coefficient of skewness given by:

$$karl - person' scofficient\ of\ skewness = \frac{Mean - Mode}{S.D}$$

In case of mode is ill – defined, the coefficient can be determined by the formula:

$$co\ efficient\ of\ skewness = \frac{3(Mean - Median)}{S.D}$$

Example 1:

Calculate Karl–Pearson's coefficient of skewness for the following data.

25, 15, 23, 40, 27, 25, 23, 25, 20

Solution:

Computation of Mean and Standard deviation:

Short – cut method.

Size **X**	**Deviation from A=25** **X-A**	$\mathbf{d^2}$
25	0	0
15	-10	100
23	-2	4
40	15	225
27	2	4
25	0	0
23	-2	4
25	0	0
20	-5	25
N = 9	$\sum$**d = – 2**	$\sum \mathbf{d^2}$ **= 362**

Mean= $A \pm \frac{\sum d}{n}$

Mean= $25 - \frac{2}{9}$

Mean= $25 - 0.22$

Mean= 24.78

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

$$\sigma = \sqrt{\frac{362}{9} - \left(\frac{-2}{9}\right)^2}$$

$$\sigma = \sqrt{40.22 - 0.05}$$

$$\sigma = \sqrt{40.17}$$

σ = 6.3

Mode = 25, as this size of item repeats 3 times

Karl – Pearson's coefficient of skewness

$$karl - person' s cofficient\ of\ skewness = \frac{Mean - Mode}{S.D}$$

$$karl - person' s cofficient\ of\ skewness = \frac{24.78 - 25}{6.3}$$

$$karl - person' s cofficient\ of\ skewness = \frac{-0.22}{6.3}$$

$karl - person' s cofficient\ of\ skewness =$-0.03

Example 2:

Find the coefficient of skewness from the data given below

Size :	3	4	5	6	7	8	9	10
Frequency :	7	10	14	35	102	136	43	8

Solution :

Size	**Frequency (f)**	**d=X-A d=X-6**	**d^2**	**fd**	**fd^2**
3	7	-3	9	-21	63
4	10	-2	4	-20	40
5	14	-1	1	-14	14
6	35	0	0	0	0
7	102	1	1	102	102
8	136	2	4	272	544
9	43	3	9	129	387
10	8	4	16	32	128
	N = 355			**$\sum fd = 480$**	**$\sum fd^2 = 1278$**

$\bar{X} = A \pm \frac{\Sigma fd}{N}$

$\bar{X} = 6 + \frac{480}{355}$

$\bar{X}$ = 6+1.35

$\bar{X}$ = 7.35

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

$\sigma = \sqrt{\frac{1278}{355} - \left(\frac{480}{355}\right)^2}$

$\sigma = \sqrt{3.6 - 1.82}$

$\sigma = \sqrt{1.78}$

σ = 1.33

Example 3:

Find Karl – Pearson's coefficient of skewness for the given distribution:

X :	0-5	5-10	10-15	15-20	**20-25**	25-30	30-35	35-40
F :	2	5	7	13	**21**	16	8	3

Solution :

The highest frequency is 21 and corresponding class interval is 20 – 25, which is the modal class.

l = Lower limit of the model class 20

f_1 = frequency of the modal class 21

f_0 = frequency of the class preceding the modal class 13

f_2 = frequency of the class succeeding the modal class 16

$\Delta_1 = f_1 - f_0$ 21-13= 8

$\Delta_2 = f_1 - f_2$ 21-16 =5

C=5

Mode = $M_{0=}$ $20 + \frac{8}{8+5} \times 5$

Mode = $M_{0=}$ $20 + \frac{8\times5}{8+5}$

Mode = $M_{0=}$ $20 + \frac{40}{13}$

Mode = $M_{0=}$ $20 + 3.076$

Mode = $M_{0=}$ 23.08

Computation of Mean and Standard deviation

X	**Mid-point M**	**Frequency f**	**Deviations $d' = \frac{M - 22.5}{5}$**	**fd'**	**d'^2**	**fd'^2**
0-5	2.5	2	-4	-8	16	32
5-10	7.5	5	-3	-15	9	45
10-15	12.5	7	-2	-14	4	28
15-20	17.5	13	-1	-13	1	13
20-25	**22.5**	21	0	0	0	0
25-30	27.5	16	1	16	1	16
30-35	32.5	8	2	16	4	32
35-40	37.5	3	3	9	9	27
		N = 75		**$\sum fd' = -9$**		**$\sum fd'^2 = 193$**

$\bar{X} = A \pm \frac{\sum fd'}{N} \times c$

$\bar{X} = 22.5 - \frac{9}{75} \times 5$

$\bar{X} = 22.5 - \frac{45}{75}$

$\bar{X} = 22.5 - 0.6$

$\bar{X} = 21.9$

$$\sigma = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} \times C$$

$$\sigma = \sqrt{\frac{193}{75} - \left(\frac{-9}{75}\right)^2} \times 5$$

$$\sigma = \sqrt{2.75 - 0.0144} \times 5$$

$$\sigma = \sqrt{2.5556} \times 5$$

$$\sigma = 1.5986 \times 5$$

$$\sigma = 7.99$$

$$karl - person' s cofficient\ of\ skewness = \frac{Mean - Mode}{S.D}$$

$$karl - person' s cofficient\ of\ skewness = \frac{21.9 - 23.08}{7.99}$$

$$karl - person' s cofficient\ of\ skewness = \frac{-1.18}{7.99} = -0.1477$$

Bowley's Coefficient of skewness:

In Karl – Pearson's method of measuring skewness the whole of the series is needed. Prof. Bowley has suggested a formula based on relative position of quartiles. In a symmetrical distribution, the quartiles are equidistant from the value of the median; i.e., Median – Q_1 = Q_3 – Median. But in a skewed distribution, the quartiles will not be equidistant from the median. Hence Bowley has suggested the following formula:

$$\text{Bowley' s Coefficient of skewness (sk)} = \frac{Q_3 + Q_1 - 2\ \text{median}}{Q_3 - Q_1}$$

Example 4:

Find the Bowley's coefficient of skewness for the following series.

2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22

Solution:

The given data in order

Position of Q_1 is = $\frac{N+1}{4}^{th}$ item

= $\frac{11+1}{4}^{th}$ item

$= \frac{12^{th}}{4}$ item

= size of 3^{th} item = 6

Position of Q3 is = $3\left(\frac{N+1}{4}\right)$

= $3\left(\frac{11+1}{4}\right)$

= size of 9^{th} item

= 18

Median (m) = $\frac{n+1}{2}$ th item

Median (m) = $\frac{11+1}{2}$ th item

= size of 6^{th} item

=12

Bowley' s Coefficient of skewness (sk) $= \frac{Q_3 + Q_1 - 2\text{ median}}{Q_3 - Q_1}$

Bowley' s Coefficient of skewness (sk) $= \frac{18+6-2(12)}{18-6}$

Bowley' s Coefficient of skewness (sk) $= \frac{24-24}{18-6}$

Bowley' s Coefficient of skewness (sk) $= \frac{0}{12} = 0$

Since sk = 0, the given series is a symmetrical data.

Exercises

1. What is coefficient of variation? What purpose does it serve?
2. What do you understand by skewness. What are the various measures of skewness.
3. Calculate Karl-Pearson's co-efficient of Skewness for the following data

Wages in Rs	10	11	12	13	14	15
Numbers	2	4	10	8	5	1

(Skewness=0.36)

4. Calculate Karl Pearson's co-efficient of skewness for the following data

Wages in Rs.	:	6	7	8	9	10	11	12
Numbers	:	3	6	9	13	8	5	4

(Skewness=0)

5. Calculate Karl-Pearson's Coefficient of skewness for the following data

Wage per item Rs.	:	12	15	20	25	30	40	50
Number of items	:	10	25	40	70	32	13	10

(Skewness=0.014)

6. Calculate Karl-Pearson's coefficient of skewness for the following data?

Size	:	9	10	11	12	13	14	15
Numbers	:	6	3	4	10	8	2	3

(Skewness=0.110)

Introduction

The term correlation is used by a common man without knowing that he is making use of the term correlation. For example when parents advice their children to work hard so that they may get good marks, they are correlating good marks with hard work.

The study related to the characteristics of only variable such as height, weight, ages, marks, wages, etc., is known as univariate analysis. The statistical Analysis related to the study of the relationship between two variables is known as Bi-Variate Analysis. Some times the variables may be inter-related. In health sciences we study the relationship between blood pressure and age, consumption level of some nutrient and weight gain, total income and medical expenditure, etc., The nature and strength of relationship may be examined by correlation and Regression analysis.

Thus Correlation refers to the relationship of two variables or more. (e-g) relation between height of father and son, yield and rainfall, wage and price index, share and debentures etc.

Correlation is statistical Analysis which measures and analyses the degree or extent to which the two variables fluctuate with reference to each other. The word relationship is important. It indicates that there is some connection between the variables. It measures the closeness of the relationship. Correlation does not indicate cause and effect relationship. Price and supply, income and expenditure are correlated.

Definitions

1. Correlation Analysis attempts to determine the degree of relationship between variables-

Ya-Kun-Chou.

2. Correlation is an analysis of the co-variation between two or more variables.- ***A.M.Tuttle.***

Correlation expresses the inter-dependence of two sets of variables upon each other. One variable may be called as (subject) independent and the other relative variable (dependent). Relative variable is measured in terms of subject.

Uses of correlation

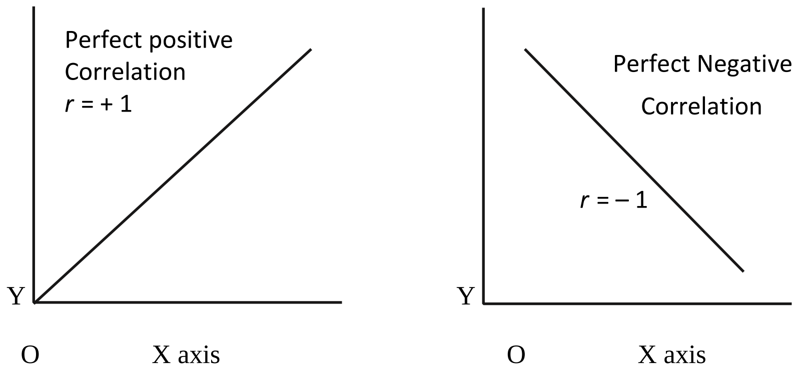
1. It is used in physical and social sciences.
2. It is useful for economists to study the relationship between variables like price, quantity etc. Businessmen estimates costs, sales, price etc. using correlation.
3. It is helpful in measuring the degree of relationship between the variables like income and expenditure, price and supply, supply and demand etc.
4. Sampling error can be calculated.
5. It is the basis for the concept of regression.

Scatter Diagram:

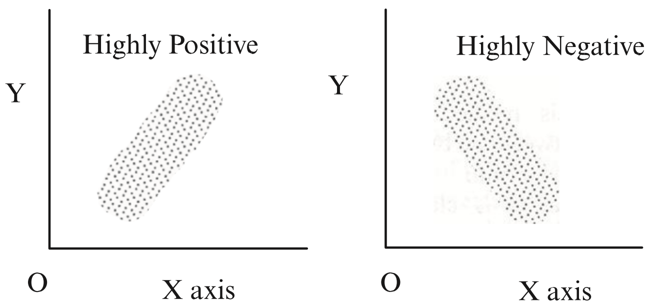
It is the simplest method of studying the relationship between two variables diagrammatically. One variable is represented along the horizontal axis and the second variable along the vertical axis. For each pair of observations of two variables, we put a dot in the plane.

There are as many dots in the plane as the number of paired observations of two variables. The direction of dots shows the scatter or concentration of various points. This will show the type of correlation.

1. If all the plotted points form a straight line from lower left hand corner to the upper right hand corner then there is Perfect positive correlation. We denote this as $r = +1$

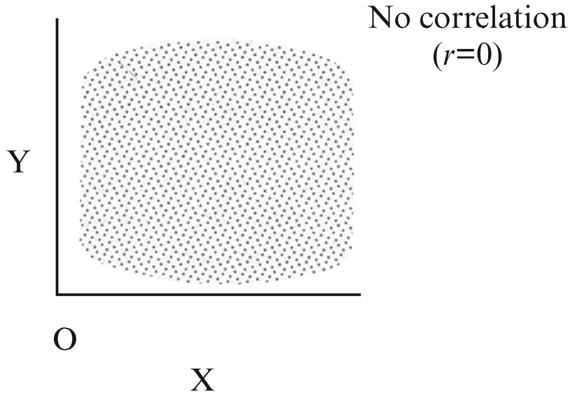


2. If all the plotted dots lie on a straight line falling from upper left hand corner to lower right hand corner, there is a perfect negative correlation between the two variables. In this case the coefficient of correlation takes the value $r = -1$.
3. If the plotted points in the plane form a band and they show a rising trend from the lower left hand corner to the upper right hand corner the two variables are highly positively correlated.



1. If the points fall in a narrow band from the upper left hand corner to the lower right hand corner, there will be a high degree of negative correlation.

2. If the plotted points in the plane are spread all over the diagram there is no correlation between the two variables.



Merits:

1. It is a simplest and attractive method of finding the nature of correlation between the two variables.
2. It is a non-mathematical method of studying correlation. It is easy to understand.
3. It is not affected by extreme items.
4. It is the first step in finding out the relation between the two variables.
5. We can have a rough idea at a glance whether it is a positive correlation or negative correlation.

Demerits:

By this method we cannot get the exact degree or correlation between the two variables.

Types of Correlation

Correlation is classified into various types. The most important ones are

i) Positive and negative.

ii) Linear and non-linear.

iii) Partial and total.

iv) Simple and Multiple.

Positive and Negative Correlation:

It depends upon the direction of change of the variables. If the two variables tend to move together in the same direction (ie) an increase in the value of one variable is accompanied by an increase in the value of the other, (or) a decrease in the value of one variable is accompanied by a decrease in the value of other, then the correlation is called positive or direct correlation. Price and supply, height and weight, yield and rainfall, are some examples of positive correlation.

If the two variables tend to move together in opposite directions so that increase (or) decrease in the value of one variable is accompanied by a decrease or increase in the value of the other variable, then the correlation is called negative (or) inverse correlation. Price and demand, yield of crop and price, are examples of negative correlation.

Linear and Non-linear correlation:

If the ratio of change between the two variables is a constant then there will be linear correlation between them.

Consider the following.

X	2	4	6	8	10	12
Y	3	6	9	12	15	18

Here the ratio of change between the two variables is the same. If we plot these points on a graph we get a straight line.

If the amount of change in one variable does not bear a constant ratio of the amount of change in the other. Then the relation is called Curvi-linear (or) non-linear correlation. The graph will be a curve.

Simple and Multiple correlation:

When we study only two variables, the relationship is simple correlation. For example, quantity of money and price level, demand and price. But in a multiple correlation we study more than two variables simultaneously. The relationship of price, demand and supply of a commodity are an example for multiple correlation.

Partial and total correlation:

The study of two variables excluding some other variable is called **Partial correlation**. For example, we study price and demand eliminating supply side. In total correlation all facts are taken into account.

Computation of correlation:

When there exists some relationship between two variables, we have to measure the degree of relationship. This measure is called the measure of correlation (or) correlation coefficient and it is denoted by **' r' .**

Algebraic or Mathematical Methods: Some of the methods of calculation of correlation coefficient are based on algebraic or mathematical treatment. The value of the coefficient of correlation by these formulae too remains between ±1. Following are the main mathematical methods –

1. Karl Pearson's Covariance method

2. Rank Correlation method

Co-variation:

The covariation between the variables x and y is defined as

Cov (x,y) = $\frac{\sum(X-\bar{X})\,(Y-\bar{Y})}{N}$

where $\bar{X}$, $\bar{Y}$ are respectively means of x and y and 'n' is the number of pairs of observations.

Karl pearson's coefficient of correlation

Karl pearson, a great biometrician and statistician, suggested a mathematical method for measuring the magnitude of linear relationship between the two variables. It is most widely used method in practice and it is known as pearsonian coefficient of correlation. It is denoted by ' r' . The formula for calculating 'r' is

(i) $r = \frac{C\,OV(x.y)}{\sigma_x.\sigma_y}$ where σ_x, σ_y are S.D of x and y

(ii) $r = \frac{\sum xy}{n\,\sigma_x\ \sigma_y}$

(iii) $r = \frac{\sum XY}{\sqrt{\sum X^2\ .\sum Y^2}}$ $X = x - \bar{x}, Y = y - \bar{y}$

When the deviations are taken from the actual mean we can apply any one of these methods. Simple formula is the third one.

The third formula is easy to calculate, and it is not necessary to calculate the standard deviations of x and y series respectively.

Steps:

1. Find the mean of the two series x and y.
2. Take deviations of the two series from x and y.

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

3. Square the deviations and get the total, of the respective squares of deviations of x and y and denote by ΣX^2, ΣY^2 respectively.
4. Multiply the deviations of x and y and get the total and Divide by n. This is covariance.
5. Substitute the values in the formula.

$$r = \frac{C\ OV(x.y)}{\sigma_x . \sigma_y} = \frac{\sum(x-\bar{x})\ \ (y-y)/n}{\sqrt{\frac{\sum(x-\bar{x})}{n}} \cdot \sqrt{\frac{\sum(y-\bar{y})}{n}}}$$

The above formula is simplified as follows

$$r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}} \qquad X = x - \bar{x}, Y = y - \bar{y}$$

Example 1:

Find Karl Pearson's coefficient of correlation from the following data between height of father (x) and son (y).

X	64	65	66	67	68	69	70
Y	66	67	65	68	70	68	72

Comment on the result.

Solution

X	**Y**	$\boldsymbol{X = x - \bar{x}}$ **X=x-67**	**X^2**	**Y=y-$\bar{y}$** **Y=y-68**	**Y^2**	**XY**
64	66	-3	9	-2	4	6
65	67	-2	4	-1	1	2
66	65	-1	1	-3	9	3

67	68	0	0	0	0	0
68	70	1	1	2	4	2
69	68	2	4	0	0	0
70	72	3	9	4	16	12
∑x=469	**∑y=476**	**0**	**∑ X$^{2=}$28**	**0**	**∑ Y$^{2=}$34**	**∑ XY =25**

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{y} = \frac{\sum y}{n}$$

$$\bar{x} = \frac{469}{7}$$

$$\bar{y} = \frac{476}{7}$$

$\bar{x} = 67$

$\bar{y} = 68$

$$r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}}$$

$$r = \frac{25}{\sqrt{28 * 34}}$$

$$r = \frac{25}{\sqrt{952}}$$

$$r = \frac{25}{30.85} = 0.81$$

Since r = + 0.81, the variables are highly positively correlated. (ie) Tall fathers have tall sons.

Example 2

From the following data compute the co-efficient of correlation between X and Y:

	X Series	**Y Series**
No. of items	15	15
Arithmetic Mean	25	18
Square of deviations from mean	136	138

Summation of product of deviations of X and Y series from their respective Arithmetic Mean is 122X

Solution:

Denoting deviations of X and Y from the arithmetic means by x and y respectively the given data are

$\sum X^2 = 136$; $\sum Y^2 = 138$

We apply Karl Pearson's method

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \cdot \sum Y^2}}$$

$$r = \frac{122}{\sqrt{136 \times 138}}$$

$$r = \frac{122}{\sqrt{18768}}$$

$$r = \frac{122}{137}$$

r = 0.89

Short-cut Method: To avoid difficult calculations due to mean being in fraction, deviations are taken from assumed means while calculating coefficient of correlation. The formula is also modified for standard deviations because deviations are taken from assumed means. Karl Pearson's formula for short-cut method is given below:

$$r = \frac{\sum d_x d_y - \frac{(\sum d_x)(\sum d_y)}{N}}{\sqrt{\sum {d_x}^2 - \frac{(\sum d_x)^2}{N}} \cdot \sqrt{\sum {d_y}^2 - \frac{(\sum d_y)^2}{N}}}$$

Or

$$r = \frac{N \sum d_x d_y - \{(\sum d_x)(\sum d_y)\}}{\sqrt{N \sum {d_x}^2 - \frac{(\sum d_x)^2}{N}} \cdot \sqrt{N \sum {d_y}^2 - \frac{(\sum d_y)^2}{N}}}$$

Example 3. The following table gives the soil temperature and the germination time at various places. Calculate the co-efficient of correlation and interpret the value

Temperature	57	42	40	38	42	45	42	44	40	46	44	43
Germination Time	10	26	30	41	29	27	27	19	18	19	31	29

Take 44 and 26 as assumed means

Solution:

We assume temperature as x and germination time as Y.

X	(X-44) dx	dx^2	Y	(Y-26) dy	dy^2	Dxdy
57	13	169	10	-16	256	-208
42	-2	4	26	0	0	0
40	-4	16	30	+4	16	-16
38	-6	36	41	+15	225	-90
42	-2	4	29	+3	9	-6
45	+1	1	27	+1	1	+1
42	-2	4	27	+1	1	-2
44	0	0	19	-7	49	0
40	-4	16	18	-8	64	+32
46	+3	4	19	-7	49	-14
44	0	0	31	+5	25	0
43	-1	1	29	+3	9	-3
N = 12	**$\sum dx = -5$**	**$\sum dx^2 = 255$**		**$\sum dy = -6$**	**$\sum dy^2 = 704$**	**$\sum dxdy = -306$**

$$r = \frac{\sum d_x d_y - \frac{(\sum d_x)(\sum d_y)}{N}}{\sqrt{\sum d_x{}^2 - \frac{(\sum d_x)^2}{N}} \ . \ \sqrt{\sum d_y{}^2 - \frac{(\sum d_y)^2}{N}}}$$

$$r = \frac{-306 - \frac{(-5)(-6)}{12}}{\sqrt{255 - \frac{(-5)^2}{12}} \ . \ \sqrt{704 - \frac{(-6)^2}{12}}}$$

$$r = \frac{-306 - \frac{30}{12}}{\sqrt{255 - \frac{25}{12}}\ .\ \sqrt{704 - \frac{36}{12}}}$$

$$r = \frac{-306 - 2.5}{\sqrt{255 - 2.1}\ .\ \sqrt{704 - 3}}$$

$$r = \frac{-308.5}{\sqrt{252.9}\ \sqrt{701}}$$

$$r = \frac{-308.5}{15.90 \times 26.47}$$

$$r = \frac{-308.5}{420.873}$$

$$r = -0.733$$

Rank Correlation

It is studied when no assumption about the parameters of the population is made. This method is based on ranks. It is useful to study the qualitative measure of attributes like honesty, colour, beauty, intelligence, character, morality etc. The individuals in the group can be arranged in order and there on, obtaining for each individual a number showing his/her rank in the group.

This method was developed by ***Edward Spearman*** in 1904. It is defined as –

$$r = 1 - \frac{6\sum D^2}{n^3 - n}$$

r = rank correlation coefficient.

Note: Some authors use the symbol ρ for rank correlation.

ΣD2 = sum of squares of differences between the pairs of ranks.

n = number of pairs of observations.

The value of r lies between –1 and +1. If r = +1, there is complete agreement in order of ranks and the direction of ranks is also same. If r = -1, then there is complete disagreement in order of ranks and they are in opposite directions.

Computation for tied observations: There may be two or more items having equal values. In such case the same rank is to be given. The ranking is said to be tied. In such circumstances an average rank is to be given to each individual item. For example if the value so is repeated twice at the 5th rank, the common rank to be assigned to each item is $\frac{5+6}{2}$ = 5.5 which is the average of 5 and 6 given as 5.5, appeared twice.

If the ranks are tied, it is required to apply a correction factor which is $\frac{1}{12}$ (m3 – m). A slightly different formula is used when there is more than one item having the same value.

The formula is

$$r = 1 - \frac{6\,[\Sigma D^2 + \frac{1}{12}(m^3 - 3) + \frac{1}{12}(m^3 - 3) + \cdots]}{n^3 - n}$$

Where 'm' is the number of items whose ranks are common and should be repeated as many times as there are tied observations.

Example 4:

In a marketing survey the price of tea and coffee in a town based on quality was found as shown below. Could you find any relation between tea and coffee price .

Price of tea	88	90	95	70	60	75	50
Price of coffee	120	134	150	115	110	140	100

Solution:

Price of tea	Rank	Price of coffee	Rank	D	D2
88	3	120	4	1	1
90	2	134	3	1	1
95	1	150	1	0	0
70	5	115	5	0	0

60	6	110	6	0	0
75	4	140	2	2	4
50	7	100	7	0	0
					$\sum D^2 = 6$

$$r = 1 - \frac{6\sum D^2}{n^3 - n}$$

$$r = 1 - \frac{6 \times 6}{7^3 - 7}$$

$$r = 1 - \frac{36}{343 - 7}$$

$$r = 1 - \frac{36}{336}$$

$$r = 1 - 0.1071$$

$$r = 0.8929$$

Example 5:

In an evaluation of answer script the following marks are awarded by the examiners.

1st	88	95	70	60	50	80	75	85
2nd	84	90	88	55	48	85	82	72

Do you agree the evaluation by the two examiners is fair?

Solution :

X	R1	y	R2	D	D2
88	2	84	4	2	4
95	1	90	1	0	0
70	6	88	2	4	16
60	7	55	7	0	0

50	8	48	8	0	0
80	4	85	3	1	1
85	3	75	6	3	9
					$\sum D^2$ =30

$$r = 1 - \frac{6\sum D^2}{n^3 - n}$$

$$r = 1 - \frac{6 \times 30}{8^3 - 8}$$

$$r = 1 - \frac{180}{512 - 8}$$

$$r = 1 - \frac{180}{504}$$

$$r = 1 - 0.357$$

$$r = 0.643$$

Example 6:

Rank Correlation for tied observations. Following are the marks obtained by 10 students in a class in two tests.

Students	A	B	C	D	E	F	G	H	I	J
Test 1	70	68	67	55	60	60	75	63	60	72
Test 2	65	65	80	60	68	58	75	63	60	70

Calculate the rank correlation coefficient between the marks of two tests.

Student	Test 1	R1	Test 2	R2	D	D2
A	70	3	65	5.5	- 2.5	6.25
B	68	4	65	5.5	-1.5	2.25

C	67	5	80	1.0	4.0	16.00
D	55	10	60	8.5	1.5	2.25
E	60	8	68	4.0	4.0	16.00
F	60	8	58	10.0	-2.0	4.00
G	75	1	75	2.0	-1.0	1.00
H	63	6	62	7.0	-1.0	1.00
I	60	8	60	8.5	0.5	0.25
J	72	2	70	3.0	-1.0	1.00
						$\sum D^2$ =50.00

60 is repeated 3 times in test 1

60, 65 is repeated twice in test 2

m=3; m=2; m=2

$$r = 1 - \frac{6\,[\sum D^2 + \frac{1}{12}(m^3 - 3) + \frac{1}{12}(m^3 - 3) + \cdots]}{n^3 - n}$$

$$r = 1 - \frac{6[50 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)]}{10^3 - 10}$$

$$r = 1 - \frac{6[50 + \frac{1}{12}(27 - 3) + \frac{1}{12}(8 - 2) + \frac{1}{12}(8 - 2)]}{1000 - 10}$$

$$r = 1 - \frac{6[50 + \frac{1}{12}(24) + \frac{1}{12}(6) + \frac{1}{12}(6)]}{990}$$

$$r = 1 - \frac{6[50 + \frac{24}{12} + \frac{6}{12} + \frac{6}{12}]}{990}$$

$$r = 1 - \frac{6[50 + 2 + 0.5 + 0.5]}{990}$$

$r = 1 - \frac{6[53]}{990}$

$r = 1 - \frac{672}{990}$

$r = 1 - 0.678$

$r = 0.322$

Questions

1. What is correlation?
2. Define Karl Pearson' s coefficient of correlation
3. What is Rank correlation? What are its merits and demerits?
4. Explain different types of correlation with examples.
5. Find Karl Pearson's co-efficient of correlation from the following data

x	10	12	18	24	23	27
y	13	18	12	25	30	10

(r=0.255)

6. Find Karl Pearson's co-efficient of correlation from the following data

X	23	27	28	29	30	31	33	35	36	39
Y	18	22	23	24	25	26	28	29	30	32

(r=0.995)

7. Find Karl pearson's co-efficient of correlation from the following data

X	:	65	66	67	67	68	69	70	72
Y	:	67	68	65	68	72	72	69	71

(r=0.603)

8. Find Karl Peason's Co-efficient of correlation from the following data

Price	:	10	12	14	16	18	20	22	24
Demand	:	20	18	15	13	12	9	10	7

(r= -0.978)

9. Rank Correlation for tied observations. Following are the marks obtained by 10 students in a class in two tests.

Students	A	B	C	D	E	F	G	H	I	J
Test 1	4	8	6	7	5	3	2	1	9	10
Test 2	4	3	7	9	10	6	5	2	1	8

Calculate the rank correlation coefficient between the marks of two tests.

(r= - 0.297)

10. In an evaluation of answer script the following marks are awarded by the examiners.

1st	30	50	80	40	70	100	20	10	60	90
2nd	60	40	90	80	100	20	30	100	50	70

Do you agree the evaluation by the two examiners is fair?

(r=0.139)

Introduction

After knowing the relationship between two variables we may be interested in estimating (predicting) the value of one variable given the value of another. The variable predicted on the basis of other variables is called the "dependent" or the 'explained' variable and the other the 'independent' or the 'predicting' variable. The prediction is based on average relationship derived statistically by regression analysis. The equation, linear or otherwise, is called the regression equation or the explaining equation.

For example, if we know that advertising and sales are correlated we may find out expected amount of sales for a given advertising expenditure or the required amount of expenditure for attaining a given amount of sales.

The relationship between two variables can be considered between, say, rainfall and agricultural production, price of an input and the overall cost of product consumer expenditure and disposable income. Thus, regression analysis reveals average relationship between two variables and this makes possible estimation or prediction.

Definition

According to Blair, "Regression is the measure of the average relationship between two or more variable in terms of the original units of the data"

According to Wallis and Robert." one of the most frequently used techniques in economics and business research, to find a relation between two or more variable that are related casually, is regression analysis"

Types of Regression

The regression analysis can be classified in to:

a) Simple and Multiple

b) Linear and Non –Linear

c) Total and Partial

a) Simple and Multiple:

In case of simple relationship only two variables are considered, for example, the influence of advertising expenditure on sales turnover. In the case of multiple relationships, more than two variables are involved. On this while one variable is a dependent variable the remaining variables are independent ones.

For example, the turnover (y) may depend on advertising expenditure (x) and the income of the people (z). Then the functional relationship can be expressed as y = f (x,z).

b) Linear and Non-linear:

The linear relationships are based on straight-line trend, the equation of which has no-power higher than one. But, remember a linear relationship can be both simple and multiple. Normally a linear relationship is taken into account because besides its simplicity, it has a better predictive value; a linear trend can be easily projected into the future. In the case of non-linear relationship curved trend lines are derived. The equations of these are parabolic.

c) Total and Partial:

In the case of total relationships all the important variables are considered. Normally, they take the form of a multiple relationships because most economic and business phenomena are affected by multiplicity of cases. In the case of partial relationship one or more variables are considered, but not all, thus excluding the influence of those not found relevant for a given purpose.

Linear Regression Equation:

If two variables have linear relationship then as the independent variable (X) changes, the dependent variable (Y) also changes. If the different values of X and Y are plotted, then the two straightlines of best fit can be made to pass through the plotted points. These two lines are known as regression lines. Again, these regression lines are based on two equations known as regression equations. These equations show best estimate of one variable for the known value of the other. The equations are linear.

Linear regression equation of **Y on X is**

$$Y = a + b X \ldots\ldots (1)$$

And X on Y is

$$X = a + b Y \ldots\ldots (2)$$

a, b are constants.

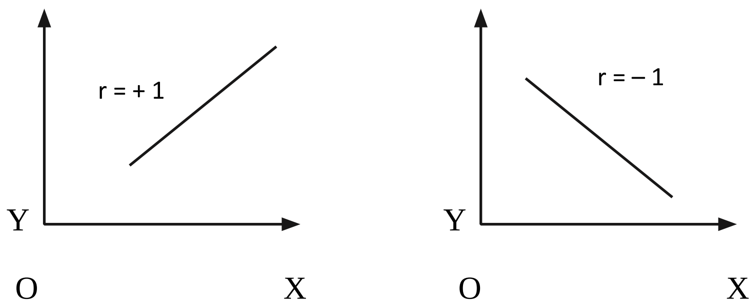
From (1) We can estimate Y for known value of X.

(2) We can estimate X for known value of Y.

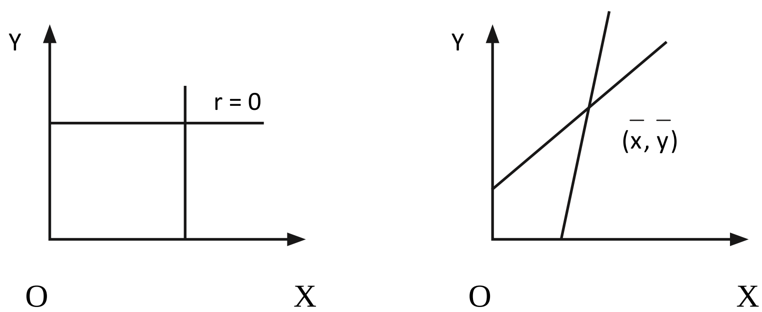
Regression Lines:

For regression analysis of two variables there are two regression lines, namely Y on X and X on Y. The two regression lines show the average relationship between the two variables.

For perfect correlation, positive or negative i.e., r = + 1, the two lines coincide i.e., we will find only one straight line. If r = 0, i.e., both the variables are independent then the two lines will cut each other at right angle. In this case the two lines will be parallel to X and Y-axes.



Lastly the two lines intersect at the point of means of X and Y. From this point of intersection, if a straight line is drawn on X axis, it will touch at the mean value of x. Similarly, a perpendicular drawn from the point of intersection of two regression lines on Y-axis will touch the mean value of Y.



Principle of 'Least Squares':

Regression shows an average relationship between two variables, which is expressed by a line of regression drawn by the method of "least squares". This line of regression can be derived graphically or algebraically. Before we discuss the various methods let us understand the meaning of least squares.

A line fitted by the method of least squares is known as the line of best fit. The line adapts to the following rules:

(i) The algebraic sum of deviation in the individual observations with reference to the regression line may be equal to zero. i.e.,

$$\sum(X - X_c) = 0 \text{ or } \sum (Y - Y_c) = 0$$

Where Xc and Yc are the values obtained by regression analysis.

(ii) The sum of the squares of these deviations is less than the sum of squares of deviations from any other line. i.e.,

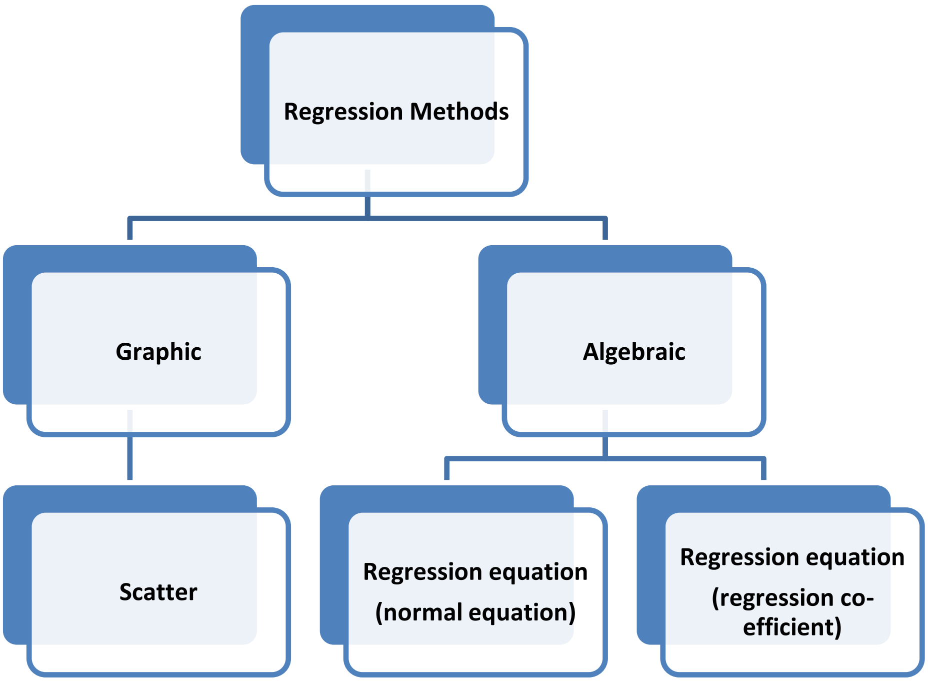
$$\sum (Y - Y_c)^2 < \sum (Y - A_i)^2$$

Where Ai = corresponding values of any other straight line.

The lines of regression (best fit) intersect at the mean values of the variables $\bar{X}$ and $\bar{Y}$, i.e., intersecting point is $\bar{x}, \bar{y}$.

Methods of Regression Analysis:

The various methods can be represented in the form of chart given below:



Graphic Method:

Scatter Diagram:

Under this method the points are plotted on a graph paper representing various parts of values of the concerned variables. These points give a picture of a scatter diagram with several points spread over. A regression line may be drawn in between these points either by free hand or by a scale rule in such a way that the squares of the vertical or the horizontal distances (as the case may be) between the points and the line of regression so drawn is the least. In other words, it should be drawn faithfully as the line of best fit leaving equal number of points on both sides in such a manner that the sum of the squares of the distances is the best.

Algebraic Methods:

(i) Regression Equation.

The two regression equations for **X on Y;**

X = a + bY

And for **Y on X; Y = a + bX**

Where X, Y are variables, and a, b are constants whose values are to be determined

For the equation, X = a + bY

The normal equations are

$\sum X = na + b \sum Y$ and

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

For the equation, Y= a + bX, the normal equations are

$\sum Y = na + b\sum X$ and

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

From these normal equations the values of *a* and *b* can be determined.

Example 1:

Find the two regression equations from the following data:

X :	6	2	10	4	8
Y :	9	11	5	8	7

Solution:

X	Y	X^2	Y^2	XY
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
$\sum X=30$	**$\sum Y=40$**	**$\sum X^2=220$**	**$\sum Y^2=340$**	**$\sum XY=214$**

Regression equation of Y on X is Y = a + bX and the normal equations are

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

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

Substituting the values, we get

40 = 5a + 30b …… (1)

214 = 30a + 220b ... (2)

Multiplying (1) by 6

240 = 30a + 180b……. (3) (2)-(3)

-26=40b

Or $b= \frac{-26}{40}$ b=-0.65

Now, substituting the value of ‘ b’ in equation (1)

40 = 5a + 30b

40=5a+30(-0.65)

40=5a-19.5

5a-19.5-40=0

$5a-59.5=0$

$5a=59.5$

$a=\frac{59.5}{5}$

$a=11.9$

Hence, required regression line Y on X is Y = 11.9 – 0.65 X.

Again, regression equation of X on Y is

$X = a + bY$ and

The normal equations are

$\sum X = na + b\sum Y$ and

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

Now, substituting the corresponding values from the above table, we get

$30 = 5a + 40b$ …. (3)

$214 = 40a + 340b$ …. (4)

Multiplying (3) by 8, we get

$240 = 40a + 320\,b$ …. (5)

(4) – (5) gives

$-26 = 20b$

$b=\frac{-26}{20}$

$b=-1.3$

Substituting b = – 1.3 in equation (3) gives

$30 = 5a - 52$

$5a = 82$

$a= \frac{82}{5}$

$a=16.4$

Hence, Required regression line of X on Y is X = 16.4 – 1.3Y

REGRESSION EQUATIONS THROUGH REGRESSION COEFFICIENTS

Regression coefficient refers to the constant value multiplied to the independent variable in a given relation. Say a relation Y = a + bx, here b (the slope of the regression line) is the sregression coefficient, since it is a multiple of independent variable x, Regression equations or lines can easily be arrived at by the use of regression coefficients. For this purpose, we are required to calculate mean, standard deviation and correlation coefficient of the given series. The following are the main methods to calculate regression coefficient Y on X (b_{yx}) or X on Y (b_{xy}).

1. Taking deviations from actual mean
2. Taking deviations from Assumed mean
3. Applying actual observations
4. Applying grouped data

Regression equation of X on Y.

This can be written as $\boldsymbol{X - \bar{X} = r\frac{\sigma x}{\sigma y}(Y - \bar{Y})}$

$\bar{X}$ is the means of X series

$\bar{Y}$ is the means of Y series

$r\frac{\sigma x}{\sigma y}$ is known as the regression co-efficient of X on Y OR (b_{xy}) = $\frac{\sum XY}{\sum Y^2}$

Regression equation of Y on X

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

$\bar{Y}$ is the means of Y series

$\bar{X}$ is the means of X series

$r\frac{\sigma y}{\sigma x}$ is known as the regression co-efficient of Y on X OR (b_{yx}) = $\frac{\sum XY}{\sum x^2}$

Example 2:

Calculate the two regression equations of X on Y and Y on X from the data given below, taking deviations from actual means of X and Y

Price (Rs)	10	12	13	12	16	15
Amount demanded	40	38	43	45	37	43

Estimate the likely demand when the price is Rs 20

Solution

X	$\boldsymbol{X-\bar{X}}$ **x-13**	$\boldsymbol{x^2}$	**Y**	$\boldsymbol{y=Y-\bar{Y}}$ **Y-41**	$\boldsymbol{y^2}$	**xy**
10	-3	9	40	-1	1	3
12	-1	1	38	-3	9	3
13	0	0	43	2	4	0
12	-1	1	45	4	16	-4
16	3	9	37	-4	16	-12
15	2	4	43	2	4	4
∑X=78	**0**	$\sum x^2 =24$	**∑Y=246**	**0**	$\sum y^2 =50$	**∑ xy =-6**

Regression equation of X on Y

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

$\bar{X}=\frac{78}{6}$ =13; $\bar{Y}=\frac{246}{6}$ =41

$r\frac{\sigma x}{\sigma y} = \frac{\sum XY}{\sum Y^2} = \frac{-6}{50}$ = -0.12

X-13 = -0.12 (y-41)

X-13= -0.12y+4.92

X=-0.12+17.92

Regression equation of Y on X

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

$\overline{X}= \frac{78}{6}$ =13; $\overline{Y}=\frac{246}{6}$ =41

$r\frac{\sigma y}{\sigma x} = \frac{\sum XY}{\sum x^2} = \frac{-6}{24}$ = - 0.25

Y-41= -0.25 (X-13)

Y-41= -0.25X+3.25

Y= - 0.25X + 44.25

Therefore

When X is 20, Y will be

Y= - 0.25X + 44.25

Y= - 0.25(20) + 44.25

Y= - 5+ 44.25

Y=39.25

When the price is Rs 20, the likely demand is **39.25**

Deviations taken from Assumed Means X and Y:

In practice we get means of fractions and for simplicity we take deviations from assumed means. When the deviations are taken from the assumed means, the procedure for finding regression equations remains the same. In case of actual means the regression equations of X on Y are

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

The value of $r\frac{\sigma x}{\sigma y}$ will now be obtained as follows

$$r\frac{\sigma x}{\sigma y} = \frac{\sum dxdy - \frac{\sum dx \times \sum dy}{N}}{\sum dy^2 - \frac{(\sum dy)^2}{N}}$$

dx= X-A; dy=Y-A

The regression equations of Y on X is

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

$$r\frac{\sigma y}{\sigma x} = \frac{\sum dxdy - \frac{\sum dx \times \sum dy}{N}}{\sum dx^2 - \frac{(\sum dx)^2}{N}}$$

Example 3:

Price indices of cotton and wool are given below for the 12 months of a year. Obtain the equations of lines of regression between the indices.

Price index of cotton (X)	78	77	85	88	87	82	81	77	76	83	97	93
Price index of wool (Y)	84	82	82	85	89	90	88	92	83	89	98	99

X	dx= X-A X- 84	dx^2	Y	Dy=Y-A Y- 88	dy^2	dx.dy
78	-6	36	84	-4	16	24
77	-7	49	82	-6	36	42
85	1	1	82	-6	36	-6
88	4	16	85	-3	9	-12
87	3	9	89	1	1	3
82	-2	4	90	2	4	-4
81	-3	9	88	0	0	0
77	-7	49	92	4	16	-28
76	-8	64	83	-5	25	40
83	-1	1	89	1	1	-1
97	13	169	98	10	100	130
93	9	81	99	11	121	99
1004	**$\sum dx = -4$**	**$\sum dx^2 = 488$**	**1061**	**$\sum dy = 5$**	**$\sum dy^2 = 365$**	**$\sum dx.dy = 287$**

Regression equation of X on Y

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

$$\boldsymbol{r\frac{\sigma x}{\sigma y} = \frac{\sum dxdy - \frac{\sum dx \quad \times \quad \sum dy}{N}}{\sum dy^2 - \frac{(\sum dy)^2}{N}}}$$

$$r\frac{\sigma x}{\sigma y} = \frac{287 - \frac{-4 \ \times \quad 5}{12}}{365 - \frac{5^2}{12}}$$

$$r\frac{\sigma x}{\sigma y} = \frac{287 \times 12 - (-20)}{365 \times 12 - 25}$$

$$r\frac{\sigma x}{\sigma y} = \frac{3444 - (-20)}{4380 - 25}$$

$$r\frac{\sigma x}{\sigma y} = \frac{3444 + 20)}{4380 - 25}$$

$$r\frac{\sigma x}{\sigma y} = \frac{3464}{4355}$$

$$r\frac{\sigma x}{\sigma y} = 0.795$$

$\bar{X}= \frac{1004}{12}$ = 83.67; $\bar{Y}=\frac{1061}{12}$ = 88.42

X-83.67 = 0.795(Y-88.42)

X-83.67= 0.795Y- 70.29

X=0.795Y-70.29+83.67

X=0.795Y+13.38

The regression equation of Y on X is

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

$$\boldsymbol{r\frac{\sigma y}{\sigma x} = \frac{\sum dxdy - \frac{\sum dx \quad \times \quad \sum dy}{N}}{\sum dx^2 - \frac{(\sum dx)^2}{N}}}$$

$$r\frac{\sigma y}{\sigma x} = \frac{287 - \frac{-4 \times 5}{12}}{488 - \frac{(-4)^2}{12}}$$

$$r\frac{\sigma y}{\sigma x} = \frac{287 \times 12 - (-20)}{488 \times 12 - 6}$$

$$r\frac{\sigma y}{\sigma x} = \frac{3444 - (-20)}{5856 - 6}$$

$$r\frac{\sigma y}{\sigma x} = \frac{3444 + 20)}{5856 - 6}$$

$$r\frac{\sigma y}{\sigma x} = \frac{3464}{5850}$$

$$r\frac{\sigma y}{\sigma x} = 0.59$$

Y-88.42 = 0.59 (X – 83.67)

Y-88.42 = 0.59X-49.37

Y= 0.59X- 49.37+88.42

Y= 0.59X+39.05

Example 4 : The following scores were worked out from a test in Mathematics and English in an annual examination.

	Scores in Mathematics (*x*)	**English (*y*)**
Mean	39.5	47.5
Standard deviation	10.8	16.8

r = + 0.42

Find both the regression equations. Using these regression estimate find the value of *Y* for *X* = 50 and the value of *X* for *Y* = 30.

Solution:

Regression of *X* on *Y*

$$\boldsymbol{X - \overline{X} = r\frac{\sigma x}{\sigma y}(Y - \overline{Y})}$$

$$\mathrm{X} - 39.5 = 0.42\ \frac{10.8}{16.8}\ (\mathrm{Y} - 47.5)$$

$$\mathrm{X} - 39.5 = 0.27\ (\mathrm{Y} - 47.5)$$

$X - 39.5 = 0.27\,Y - 12.825$

$X = 0.27\,Y - 12.825 + 39.5$

$X = 0.27\,Y + 26.675$

When Y = 30, the value of X

$X = 0.27\,(30) + 26.675$

$X = 8.1 + 26.675$

$X = 34.775$

Regression equation of Y on X

$$\boldsymbol{Y - \overline{Y} = r\frac{\sigma y}{\sigma x}(X - \overline{X})}$$

$$\boldsymbol{Y - 47.5 = 0.42\,\frac{16.8}{10.8}(X - 39.5)}$$

$$Y - 47.5 = 0.42\,\frac{16.8}{10.8}(X - 39.5)$$

$Y - 47.5 = 0.653\ \ (X - 39.5)$

$Y - 47.5 = 0.653X - 25.7935)$

$Y = 0.653X - 25.7935 + 47.5$

$Y = 0.653X + 21.7065$

When X is 50, the value of Y

$Y = 0.653(50) + 21.7065$

$Y = 32.65 + 21.7065$

$Y = 54.36$

Exercises

1. Explain the uses of regression analysis
2. Distinguish between correlation and regression.
3. Obtain the two regression equations from the following

x	3	6	5	4	4	6	7	5
y	3	2	3	5	3	6	6	4

X=0.375y+3.5, Y=1.5+0.5x

4. Obtain the two regression equations of the following

X	:	45	42	44	43	41	45	43	40
Y	:	40	38	36	35	38	39	37	41

X=-0.178y+49.66, Y=47.37-0.218x

5. Calculate the two regression equations from the following

X	:	10	12	13	12	16	15
Y	:	40	38	43	45	37	43

X=-0.12y+17.92, Y=44.25-0.25x

6. Obtain the two regression equation from the following

X	:	1	2	3	4	5
Y	:	2	3	5	4	6

X=0.9y-0.6, Y=1.3+0.9x

Introduction

Arrangement of statistical data in chronological order ie., in accordance with occurrence of time, is known as “Time Series”. Such series have a unique important place in the field of Economic and Business statistics. An economist is interested in estimating the likely population in the coming year so that proper planning can be carried out with regard to food supply, job for the people etc. Similarly, a business man is interested in finding out his likely sales in the near future, so that the businessman could adjust his production accordingly and avoid the possibility of inadequate production to meet the demand. In this connection one usually deal with statistical data, which are collected, observed or recorded at successive intervals of time. Such data are generally referred to as ‘time series’.

Definition

According to Mooris Hamburg, “A time series is a set of statistical observations arranged in chronological order”.

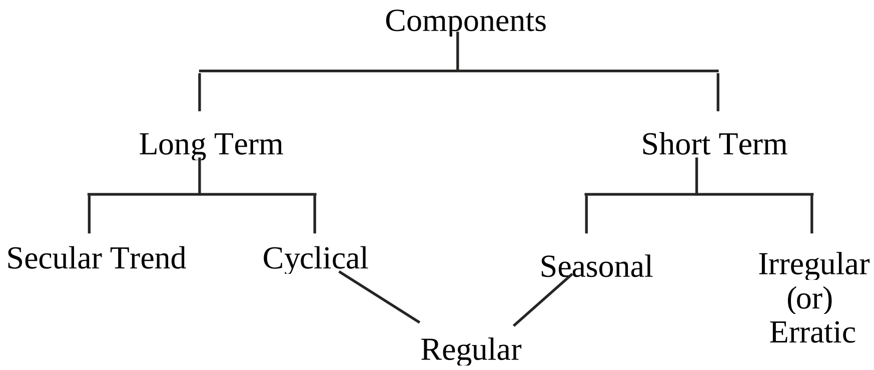
Ya-Lun- chou defining the time series as “A time series may be defined as a collection of readings belonging to different time periods, of some economic variable or composite of variables. A time series is a set of observations of a variable usually at equal intervals of time. Here time may be yearly, monthly, weekly, daily or even hourly usually at equal intervals of time.

Hourly temperature reading, daily sales, monthly production are examples of time series. Number of factors affect the observations of time series continuously, some with equal intervals of time and others are erratic studying, interpreting analyzing the factors is called Analysis of Time Series.

The Primary purpose of the analysis of time series is to discover and measure all types of variations which characterise a time series. The central objective is to decompose the various elements present in a time series and to use them in business decision making.

Components of Time series

The components of a time series are the various elements which can be segregated from the observed data. The following are the broad classification of these components.



In time series analysis, it is assumed that there is a multiplicative relationship between these four components.

Symbolically,

$$Y = T \times S \times C \times I$$

Where Y denotes the result of the four elements; T = Trend ; S = Seasonal component; C = Cyclical components; I = Irregular component

In the multiplicative model it is assumed that the four components are due to different causes but they are not necessarily independent and they can affect one another.

Another approach is to treat each observation of a time series as the sum of these four components. Symbolically

$$Y = T + S + C + I$$

The additive model assumes that all the components of the time series are independent of one another.

1) Secular Trend or Long - Term movement or simply Trend
2) Seasonal Variation
3) Cyclical Variations
4) Irregular or erratic or random movements(fluctuations)

1. Secular Trend:

It is a long term movement in Time series. The general tendency of the time series is to increase or decrease or stagnate during a long period of time is called the secular trend or simply trend. Population growth, improved technological progress, changes in consumers taste are the various factors of upward trend. We may notice downward trend relating to deaths, epidemics, due to improved medical facilities and sanitations. Thus a time series shows fluctuations in the upward or downward direction in the long run.

Methods of Measuring Trend:

Trend is measured by the following mathematical methods.

1. Graphical method
2. Method of Semi-averages
3. Method of moving averages
4. Method of Least Squares

Graphical Method:

This is the easiest and simplest method of measuring trend. In this method, given data must be plotted on the graph, taking time on the horizontal axis and values on the vertical axis. Draw a smooth curve which will show the direction of the trend. While fitting a trend line the following important points should be noted to get a perfect trend line.

(i) The curve should be smooth.

(ii) As far as possible there must be equal number of points above and below the trend line.

(iii) The sum of the squares of the vertical deviations from the trend should be as small as possible.

(iv) If there are cycles, equal number of cycles should be above or below the trend line.

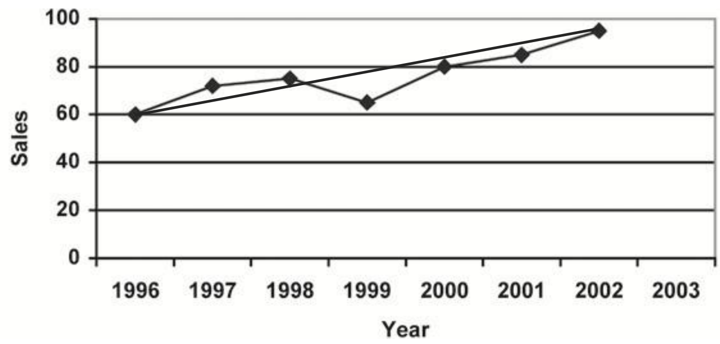
(v) In case of cyclical data, the area of the cycles above and below should be nearly equal.

Example 1 :

Fit a trend line to the following data by graphical method.

Year	1996	1997	1998	1999	2000	2001	2002
Sales (in Rs.'000)	60	72	75	65	80	85	95

Solution:



The dotted lines refers trend line

Merits:

1. It is the simplest and easiest method. It saves time and labour.
2. It can be used to describe all kinds of trends.
3. This can be used widely in application.
4. It helps to understand the character of time series and to select appropriate trend.

Demerits:

1. It is highly subjective. Different trend curves will be obtained by different persons for the same set of data.
2. It is dangerous to use freehand trend for forecasting purposes.
3. It does not enable us to measure trend in precise quantitative terms.

Method of semi averages

In this method, the given data is divided into two parts, preferably with the same number of years. For example, if we are given data from 1981 to 1998 i.e., over a period of 18 years, the two equal parts will be first nine years, i.e., 1981 to 1989 and from 1990 to 1998. In case of odd number of years like 5,7,9,11 etc, two equal parts can be made simply by omitting the middle year. For example, if the data are given for 7 years from 1991 to 1997, the two equal parts would be from 1991 to 1993 and from 1995 to 1997, the middle year 1994 will be omitted.

After the data have been divided into two parts, an average of each parts is obtained. Thus we get two points. Each point is plotted at the mid-point of the class interval covered by respective part and then the two points are joined by a straight line which gives us the required trend line. The line can be extended downwards and upwards to get intermediate values or to predict future values.

Example 2:

Draw a trend line by the method of semi-averages.

Year	1991	1992	1993	1994	1995	1996
Sales (in Rs.1000)	60	75	81	110	106	120

Solution:

Year	Sales (Rs)	Semi Total	Semi – average	Trend value
1991	60			59
1992	75	216	72	72
1993	81			85
1994	110			98
1995	106	333	111	111
1996	117			124

Difference in middle periods = 1995 –1992 = 3 years

Difference in semi averages = 111 –72 = 39

∴Annual increase in trend = 39/3 = 13

Trend of 1991 = Trend of 1992 -13

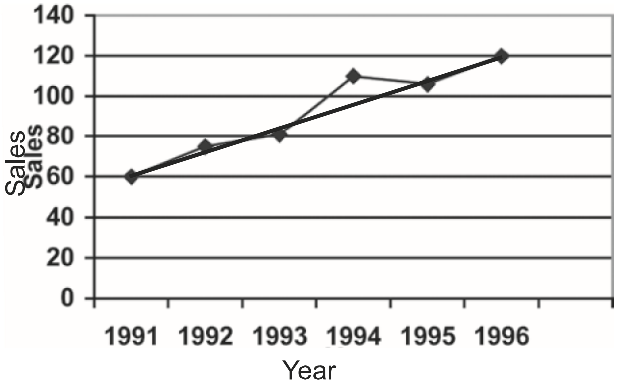
= 72 – 13 = 59

Trend of 1993 = Trend of 1992 +13

= 72 + 13 = 85

Similarly, we can find all the values

The following graph will show clearly the trend line.



Example 3:

Calculate the trend value to the following data by the method of semi- averages.

Year	1995	1996	1997	1998	1999	2000	2001
Expenditure (Rs. in Lakhs)	1.5	1.8	2.0	2.3	2.4	2.6	3.0

Solution:

Year	Expenditure (Rs.)	Semi total	Semi average	Trend values
1995	1.5			1.545
1996	1.8	5.3	1.77	1.77
1997	2			1.995
1998	2.3			2.22
1999	2.4	7.3	2.43	2.445
2000	2.6			2.67
2001	3			2.895

Difference between middle periods = 2000 – 1996

= 4 years

Difference between semi-averages = 2.67 - 1.77

= 0.9

Annual trend values = $\frac{0.9}{4}$

= 0.225

Trend of 1995 = Trend of 1996 – 0.225

= 1.77 – 0.225

= 1.545

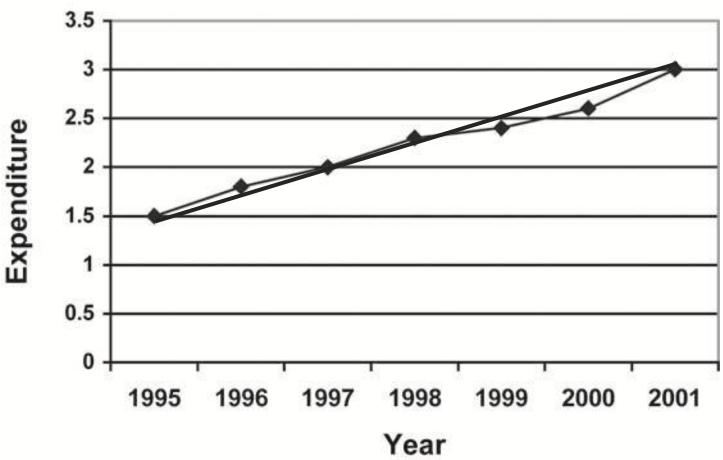
Trend of 1996 = 1.77

Trend of 1997 = 1.77 + 0.225

= 1.995

Similarly we can find all the trend values

Similarly we can find all the trend values



Merits:

1. It is simple and easy to calculate
2. By this method every one getting same trend line.

3. Since the line can be extended in both ways, we can find the later and earlier estimates.

Demerits:

1. This method assumes the presence of linear trend to the values of time series which may not exist.
2. The trend values and the predicted values obtained by this method are not very reliable.

Method of Moving Averages:

This method is very simple. It is based on Arithmetic mean. Theses means are calculated from overlapping groups of successive time series data. Each moving average is based on values covering a fixed time interval, called "period of moving average" and is shown against the center of the interval. The method of 'odd period' of moving average is as follows. (3 or 5). The moving averages for three years is $\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3}$ etc.

The formula for five yearly moving average is $\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5}$ etc.

Steps for calculating odd number of years.

1. Find the value of three years total, place the value against the second year.
2. Leave the first value and add the next three years value (ie 2nd, 3rd and 4th years value) and put it against 3rd year.
3. Continue this process until the last year' s value taken.
4. Each total is divided by three and placed in the next column.

These are the trend values by the method of moving averages

Example 4 :

Calculate the three yearly average of the following data.

Year	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
Production	50	36	43	45	39	38	33	42	41	34

Solution :

Year	Production (in tones)	3 years moving total	3 years moving average as Trend values
1975	50	-	-
1976	36	129	43.0
1977	43	124	41.3
1978	45	127	42.3
1979	39	122	40.7
1980	38	110	36.7
1981	33	113	37.7
1982	42	116	38.7
1983	41	117	39.0
1984	34	-	-

Even Period of Moving Averages:

When the moving period is even, the middle period of each set of values lies between the two time points. So we must center the moving averages.

The steps are

1. Find the total for first 4 years and place it against the middle of the 2nd and 3rd year in the third column.
2. Leave the first year value, and find the total of next four-year and place it between the 3rd and 4th year.
3. Continue this process until the last value is taken.

4. Next, compute the total of the first two four year totals and place it against the 3rd year in the fourth column.
5. Leave the first four years total and find the total of the next two four years' totals and place it against the fourth year.
6. This process is continued till the last two four years' total is taken into account.
7. Divide this total by 8 (Since it is the total of 8 years) and put it in the fifth column.

These are the trend values.

Example 5 :

The production of Tea in India is given as follows. Calculate the Four-yearly moving averages

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Production	464	515	518	467	502	540	557	571	586	612

Solution:

Year	Production (in tones)	4 years Moving total	Total of Two four years	Trend Values
1993	464		-	-
		-		
1994	515			
		1964		
1995	518		3966	495.8
		2002		
1996	467		4029	503.6
		2027		
1997	502		4093	511.6
		2066		
1998	540		4236	529.5

		2170		
1999	557		4424	553.0
		2254		
2000	571		4580	572.5
		2326		
2001	586			
		-		
2002	612			

Example

Using five year moving average determine the trend from the following data

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Production	332	317	357	392	402	405	410	427	405	431

Year	Production	5 years moving total	5 years moving average as Trend values
1991	332		
1992	317		
1993	357	1800	360
1994	392	1873	374.6
1995	402	1966	393.2
1996	405	2036	407.2
1997	410	2049	409.8
1998	427	2078	415.6
1999	405		
2000	431		

Merits:

1. The method is simple to understand and easy to adopt as compared to other methods.
2. It is very flexible in the sense that the addition of a few more figures to the data, the entire calculations are not changed. We only get some more trend values.
3. Regular cyclical variations can be completely eliminated by a period of moving average equal to the period of cycles.
4. It is particularly effective if the trend of a series is very irregular.

Demerits:

1. It cannot be used for forecasting or predicting future trend, which is the main objective of trend analysis.
2. The choice of the period of moving average is sometimes subjective.
3. Moving averages are generally affected by extreme values of items.
4. It cannot eliminate irregular variations completely.

Method of Least Square

This method is widely used. It plays an important role in finding the trend values of economic and business time series. It helps for forecasting and predicting the future values. The trend line by this method is called the line of best fit.

The equation of the trend line is $y = a + bx$, where the constants a and b are to be estimated so as to minimize the sum of the squares of the difference between the given values of y and the estimate values of y by using the equation. The constants can be obtained by solving two normal equations.

$$\sum y = na + b\sum x \text{ ………. (1)}$$

$$\sum xy = a\sum x + b\sum x^2 \text{ ……… (2)}$$

Here x represent time point and y are observed values. ‘ n’ is the number of pair-values.

When odd number of years are given

Step 1: Writing given years in column 1 and the corresponding sales or production etc in column 2.

Step 2: Write in column 3 start with 0, 1, 2 .. against column 1 and denote it as X

Step 3: Take the middle value of X as A

Step 4: Find the deviations u = X - A and write in column 4

Step 5: Find u^2 values and write in column 5.

Step 6: Column 6 gives the product *uy*

Now the normal equations become

$\sum y = na + b\sum u$ (1) where u = X-A

$\sum uy = a\sum u + b\sum u^2$ (2)

Since $\sum u = 0$, From equation (1)

$$a = \frac{\sum y}{n}$$

From equation (2)

$$\sum uy = b\sum u^2$$

Therefore $b = \frac{\sum uy}{\sum u^2}$

∴ The fitted straight line is

$$y = a + bu = a + b(X - A)$$

Example 6:

For the following data, find the trend values by using the method of Least squares

Year	1990	1991	1992	1993	1994
Production (in tones)	50	55	45	52	54

Estimate the production for the year 1996.

Solution :

Year (x)	Production (y)	X = x - 1990	u = x-A = x-2	u^2	Uy	Trend values
1990	50	0	– 2	4	– 100	50.2
1991	55	1	– 1	1	– 55	50.7
1992	45	**2 (A)**	0	0	0	51.2
1993	52	3	1	1	52	51.7
1994	54	4	2	4	108	52.2
Total	256			10	5	

Where A is an assumed value

The equation of straight line is

Y = a + bX

= a + bu , where u = X - 2

the normal equations are

$\sum y = na + b\sum u$……….(1)

$\sum uy = a\sum u + b\sum u^2$ …..(2)

since

$\sum u = 0$ from(1) $\sum y = na$

$$a = \frac{\sum y}{n}$$

$= \frac{256}{5} = 51.2$

From equation (2)

$\sum uy = b\sum u^2$

5 =10b

b= $\frac{5}{10} = 0.5$

The fitted straight line is

y = a + bu

y = 51.2 + 0.5 (X-2)

y = 51.2 + 0.5X -1.0

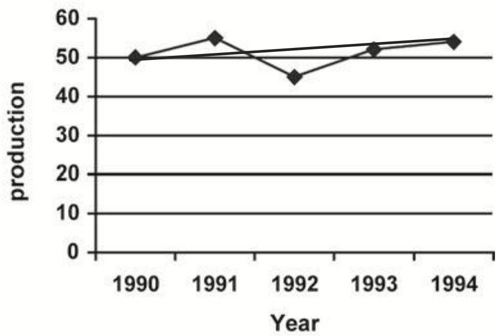
y = 50.2 + 0.5X

Trend values are, 50.2, 50.7, 51.2, 51.7, 52.2

The estimate production in 1996 is put X = x – 1990

X = 1996 –1990 = 6

Y = 50.2 + 0.5X = 50.2 +0.5(6) = 50.2 +3.0 = 53.2 tonnes.



When even number of years are given

Here we take the mean of middle two values of X as A Then u = $\frac{X-A}{1/2} = 2(X - A)$

The other steps are as given in the odd number of years.

Example 7:

Fit a straight line trend by the method of least squares for the following data.

Year	1983	1984	1985	1986	1987	1988
Sales (Rs. in lakhs)	3	8	7	9	11	14

Also estimate the sales for the year 1991

Solution:

Year (x)	Sales (y)	X = x-1983	u = 2X - 5	u^2	uy	Trend values
1983	3	0	– 5	25	–15	3.97
1984	8	1	– 3	9	– 24	5.85
1985	7	2	– 1	1	– 7	7.73
1986	9	3	1	1	9	9.61
1987	11	4	3	9	33	11.49
1988	14	5	5	25	70	13.37
Total	**52**		**0**	**70**	**66**	

u = X − A

1 2/

= 2(X −2 5.) = 2X −5

The straight line equation is

$y = a + bX = a + bu$

The normal equations are

$\sum y = na$ …….(1)

$\sum uy = b\sum u^2$ ……(2)

From (1) 52 = 6a

$a = \frac{52}{6}$

= 8 67.

From (2) 66 = 70b

$b = \frac{70}{66}$

= 0 94.

The fitted straight line equation is

$$y = a + bu$$

$$y = 8.67+0.94(2X-5)$$

$$y = 8.67 + 1.88X - 4.7$$

$$y = 3.97 + 1.88X \text{ -----------(3)}$$

The trend values are

Put $X = 0, y = 3.97$ $\quad$ $X = 1, y = 5.85$

$X = 2, y = 7.73$ $\quad$ $X = 3, y = 9.61$

$X = 4, y = 11.49$ $\quad$ $X = 5, y = 13.37$

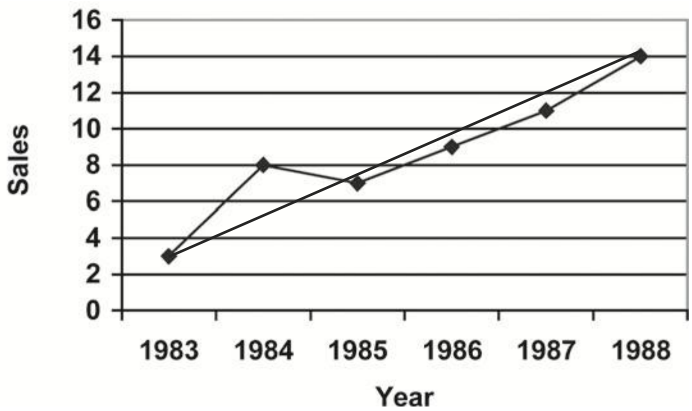
The estimated sale for the year 1991 is; put $X = x - 1983$

$$= 1991 - 1983 = 8$$

$$y = 3.97 + 1.88 \times 8$$

$$= 19.01 \text{ lakhs}$$

The following graph will show clearly the trend line.



Merits:

1. Since it is a mathematical method, it is not subjective so it eliminates personal bias of the investigator.
2. By this method we can estimate the future values. As well as intermediate values of the time series.
3. By this method we can find all the trend values.

Demerits:

1. It is a difficult method. Addition of new observations makes recalculations.
2. Assumption of straight line may sometimes be misleading since economics and business time series are not linear.
3. It ignores cyclical, seasonal and irregular fluctuations.
4. The trend can estimate only for immediate future and not for distant future.

Seasonal Variations:

Seasonal Variations are fluctuations within a year during the season. The factors that cause seasonal variation are

i) Climate and weather condition.

ii) Customs and traditional habits.

For example the sale of ice-creams increase in summer, the umbrella sales increase in rainy season, sales of woolen clothes increase in winter season and agricultural production depends upon the monsoon etc.,

Secondly in marriage season the price of gold will increase, sale of crackers and new clothes increase in festival times.

So seasonal variations are of great importance to businessmen, producers and sellers for planning the future. The main objective of the measurement of seasonal variations is to study their effect and isolate them from the trend.

Measurement of seasonal variation:

The following are some of the methods more popularly used for measuring the seasonal variations.

1. Method of simple averages.
2. Ratio to trend method.
3. Ratio to moving average method.
4. Link relative method

Among the above four methods the method of simple averages is easy to compute seasonal variations.

Method of simple averages

The steps for calculations:

i) Arrange the data season wise

ii) Compute the average for each season.

iii) Calculate the grand average, which is the average of seasonal averages.

iv) Obtain the seasonal indices by expressing each season as percentage of Grand average

The total of these indices would be 100n where ' n' is the number of seasons in the year.

Example 8:

Find the seasonal variations by simple average method for the data given below.

Quarter

Year	I	II	III	IV
1989	30	40	36	34
1990	34	52	50	44
1991	40	58	54	48
1992	54	76	68	62
1993	80	92	86	82

Solution :

Quarter

Year	**I**	**II**	**III**	**IV**
1989	30	40	36	34
1990	34	52	50	44

1991	40	58	54	48
1992	54	76	68	62
1993	80	92	86	82
Total	238	318	294	270
Average	47.6	63.6	58.8	54
Seasonal Indices	**85**	**113.6**	**105**	**96.4**

Grand average = $\frac{47.6+63.6+58.8+54}{4}$

$= \frac{224}{4} = 56$

Seasonal index for

I Quarter = $\frac{first\ quartely\ average}{grand\ average} \times 100$

$= \frac{47.6}{56} \times 100 = 85$

Seasonal index for II Quarter = $\frac{Second\ quartely\ average}{grand\ average} \times 100$

$= \frac{63.6}{56} \times 100 = 113.6$

Seasonal index for III Quarter = $\frac{Third\ quartely\ average}{grand\ average} \times 100$

$= \frac{58.5}{56} \times 100 = 105$

Seasonal index for IV Quarter = $\frac{fourth\ quartely\ average}{grand\ average}$

$= \frac{54}{56} \times 100 = 96.4$

Example 9:

Calculate the seasonal indices from the following data using simple average method.

Year

Quarter	1974	1975	1976	1977	1978
I	72	76	74	76	74
II	68	70	66	74	74
III	80	82	84	84	86
IV	70	74	80	78	82

Solution :

Quarter

Year	**I**	**II**	**III**	**IV**
1974	72	68	80	70
1975	76	70	82	74
1976	74	66	84	80
1977	76	74	84	78
1978	74	74	86	82
Total	372	352	416	384
Average	74.45	70.4	83.2	76.8
Seasonal Indices	**97.6**	**92.4**	**109.2**	**100.8**

Grand average = $\frac{74.4+70.4+83.2+76.8}{4}$

$= \frac{304.8}{4} = 76.2$

Seasonal index for I Quarter = $\frac{first\ quartely\ average}{grand\ average} \times 100$

$= \frac{74.4}{76.20} \times 100 = 97.6$

Seasonal index for II Quarter = $\frac{Second\ quartely\ average}{grand\ average} \times 100$

$= \frac{70.4}{76.2} \times 100 = 92.4$

Seasonal index for III Quarter = $\frac{Third\ quartely\ average}{grand\ average} \times 100$

$= \frac{83.2}{76.2} \times 100 = 109.2$

Seasonal index for IV Quarter = $\frac{fourth\ \ quartely\ average}{grand\ average}$

$= \frac{76.8}{76.2} \times 100 = 100.8$

The total of seasonal indices calculated must be equal to 400 here we have

= 97.6 + 92.4 + 109.2 + 100.8

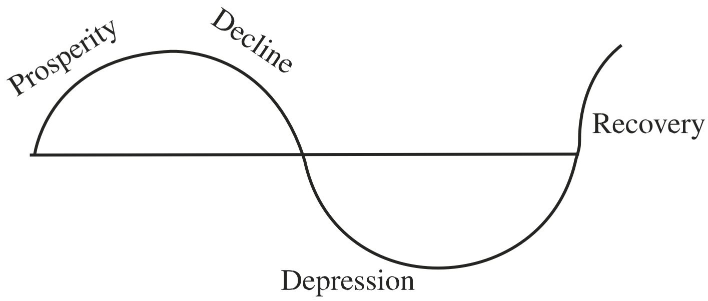
= 400 hence verified.

Cyclical variations:

The term cycle refers to the recurrent variations in time series, that extend over longer period of time, usually two or more years. Most of the time series relating to economic and business show some kind of cyclic variation. A business cycle consists of the recurrence of the up and down movement of business activity. It is a four-phase cycle namely.

1. Prosperity 2. Decline 3. Depression 4. Recovery

Each phase changes gradually into the following phase. The following diagram illustrates a business cycle.



The study of cyclical variation is extremely useful in framing suitable policies for stabilising the level of business activities. Businessmen can take timely steps in maintaining business during booms and depression.

Irregular variation:

Irregular variations are also called erratic. These variations are not regular and which do not repeat in a definite pattern.

These variations are caused by war, earthquakes, strikes flood, revolution etc. This variation is short-term one, but it affect all the components of series. There is no statistical techniques for measuring or isolating erratic fluctuation. Therefore the residual that remains after eliminating systematic components is taken as representing irregular variations.

1. Explain about seasonal variation
2. Give the names of different methods of measuring trend.
3. What are the merits and demerits of the semi-average method?
4. What are the components of time series.
5. Fit a straight line trend by the method of least square and estimate the production in 1990

Year	:	1981	1982	1983	1984	1985
Production in quantities	:	15	16.5	18	20.5	25

Y=19+2.4x , Production for 1990 = 35.8

6. Using five year moving average method determine the trend from the following

Year	:	1986	1987	1988	1989	1990	1991	1992	1993	1994
Expense	:	50	36	43	44	38	38	32	38	41

(42.2, 39.8, 39, 38, 37.4)

7. Fit a straight line trend by the method of least squares and estimate the net profit in 2012. What is the amount of change per annum in net profit?

Year	:	2004	2005	2006	2007	2008	2009	2010
Net Profit (Rs.Crores)	:	32	36	44	37	71	72	109

Y=21.92+11.79x

Profit for 2012=116.24

8. Find out trend values by the method of moving average (5 yearly)

Year	:	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Production	:	351	366	361	362	400	419	410	420	450	500	518

(368,381.6, 390.4, 402.2, 419.8, 439.8, 459.6)

9. Fit a straight line trend by the method of least square to the following data. Also estimate the value for 2004.

Year	:	1997	1998	1999	2000	2001	2002	2003
Production of steel in tons	:	60	72	75	65	80	85	95

Y=61.429+4.857x

Production for 2004=95.428

10. Find out trend value by the method of four yearly moving average for the following data

Year	:	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
Production	:	37.4	31.1	38.7	39.5	47.9	42.6	48.4	64.6	58.4	38.6	51.4	84.4

(37.98, 40.74, 43.39, 47.74, 52.19, 53, 52.88, 55.73)

11. Fit a straight line trend by the method of least square and estimate the net profit in 2014.

Year	2006	2007	2008	2009	2010	2011	2012
Net Profit in Rs. Crores	35	40	38	44	37	100	89

Y=24.605+10.035x

Profit for 2014=104.885

12. Using five year moving average determine the trend from the following data

Year	1970	1971	1972	1973	1974	197	1976
Income	1.5	1.8	2.0	2.3	2.4	2.6	3.0

(2, 2.22, 2.46)

Introduction

The word "Geometry" is derived from the Greek word "geo" meaning "earth" and "metron" meaning "measuring". The need of measuring land is the origin of geometry.

The branch of mathematics where algebraic methods are employed for solving problem in geometry is known as Analytical Geometry. It is sometimes called cartesian Geometry after the french mathematician Des-Cartes.

Section formula

Internal division

If the point P divides the line segment joining two points A(x_1, x_2) and B (y_1, y_2) internally in the ratio of *m:n* is called internal division. The formula for internal division is

$$(\boldsymbol{x}, \boldsymbol{y}) = \left(\frac{\boldsymbol{m x_2 + n x_1}}{\boldsymbol{m + n}}, \frac{\boldsymbol{m y_2 + n y_1}}{\boldsymbol{m + n}}\right)$$

External division

If the point P divides the line segment joining two points A(x_1, x_2) and B (y_1, y_2) externally in the ratio of *m:n* is called external division. The formula for internal division is

$$(\boldsymbol{x}, \boldsymbol{y}) = \left(\frac{\boldsymbol{m x_2 - n x_1}}{\boldsymbol{m - n}}, \frac{\boldsymbol{m y_2 - n y_1}}{\boldsymbol{m - n}}\right)$$

Example 1

Find the point which divides the line segment joining the points (3 , 5) and (8 , 10) internally in the ratio 2 : 3.

Solution Let *A* (3 ,5) and *B* (8 ,10) be the given points.

Let the point *P*(*x*,*y*) divide the line AB internally in the ratio 2 :3.

By section formula,

$$p(x, y) = \left(\frac{m x_2 + n x_1}{m + n}, \frac{m y_2 + n y_1}{m + n}\right)$$

$$p(x, y) = \left(\frac{2(8) + 3(3)}{2 + 3}, \frac{2(10) + 3(5)}{2 + 3}\right)$$

$$p(x,y) = \left(\frac{16+9}{5}, \frac{20+15}{5}\right)$$

$$p(x,y) = \left(\frac{25}{5}, \frac{35}{5}\right)$$

$$p(x,y) = (5,7)$$

Example 2

Find the point which divides the line segment joining the points (2 , 1) and (3 , 5) externally in the ratio 2 : 3.

Solution Let *A* (2,1) and *B* (3,5) be the given points.

Let the point *P*(*x*,*y*) divide the line AB externally in the ratio 2 :3.

By section formula,

$$(x,y) = \left(\frac{mx_2 - nx_1}{m-n}, \frac{my_2 - ny_1}{m-n}\right)$$

$$p(x,y) = \left(\frac{2(3)-3(2)}{2-3}, \frac{2(5)-3(1)}{2-3}\right)$$

$$p(x,y) = \left(\frac{6-6}{-1}, \frac{10-3}{-1}\right)$$

$$p(x,y) = \left(\frac{0}{-1}, \frac{7}{-1}\right)$$

$$p(x,y) = (0,-7)$$

Midpoint

In geometry, the **midpoint** is the middle point of a line segment. It is equidistant from both endpoints, and it is the centriod both of the segment and of the endpoints. It bisects the segment.

$$\mathbf{(x,y)} = \left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$$

Example 3

Find the midpoint of the line segment joining the points (3,0) and (-1,4)

Solution

Midpoint $M(x , y)$ of the line segment joining the points $(x_1 y_1)$ and $(x_2 y_2)$

M(x,y) = $\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$

∴ Mid- point of the line segment joining the points (3, 0) and (-1, 4) is

M(x,y) = $\left(\frac{3+(-1)}{2}, \frac{0+4}{2}\right)$

M(x,y) = $\left(\frac{2}{2}, \frac{4}{2}\right) =$ M(1,2)

Centroid of a triangle

The centroid of the triangle whose vertices is $(x_1 y_1)$ $(x_2 y_2)$ and $(x_3 y_3)$ is

(x,y)= $\left(\boldsymbol{\frac{x_1+x_2+x_3}{3}}, \boldsymbol{\frac{y_1+y_2+y_3}{3}}\right)$

Example 4

Find the co –ordinates of the centriod of the triangle whose vertices are (3,2) , (-1,-4) and (-5,6)

(x,y)= $\left(\frac{x_1+x_2+x_3}{3}, \frac{y_1+y_2+y_3}{3}\right)$

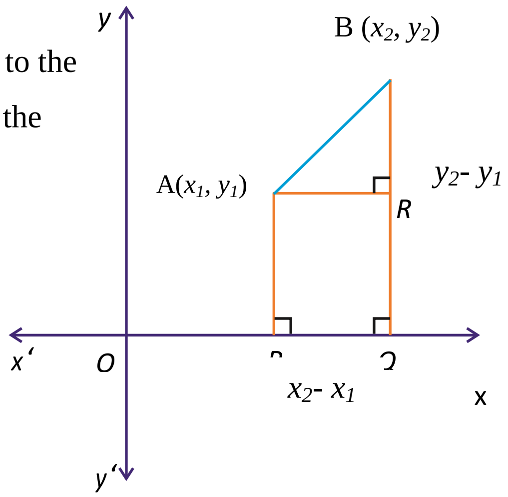
(x,y)= $\left(\frac{3+(-1)+(-5)}{3}, \frac{2+(-4)+6}{3}\right)$

(x,y)= $\left(\frac{-3}{3}, \frac{4}{3}\right)$

(x,y)= -1$\left(, \frac{4}{3}\right)$

DISTANCE BETWEEN TWO POINTS

Let $A(x_1, y_1)$ and $B(x_2, y_2)$ be any two points in the plane. We shall now find the distance between these two points.



Let P and Q be the foot of the perpendiculars from A and B to the x-axis respectively. AR is drawn perpendicular to BQ. From the diagram,

$AR = PQ = OQ - OP = x_2 - x_1$

and

$BR = BQ - RQ = y_2 - y_1$

From right angle ARB

$AB^2 = AR^2 + RB^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$

$\therefore$ **AB** = $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

Hence the distance between the points A and B is

AB = $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

Example 5

Find the distance between the points (-4, 0) and (3, 0)

Solution The points (-4, 0) and (3, 0) lie on the x-axis. Hence

d = $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

AB = $\sqrt{(3 - -4)^2 + (0 - 0)^2}$

AB = $\sqrt{(7)^2 + (0)^2}$

AB = $\sqrt{49}$

AB = 7

Example 6

Show that the three points (4, 2), (7, 5) and (9, 7) lie on a straight line.

Solution

Let the points be A (4, 2), B (7, 5) and C (9, 7). By the distance formula

$AB^2 = (4-7)^2 + (2-5)^2 = (-3)^2 + (-3)^2 = 9+9 = 18$

$BC^2 = (9-7)^2 + (7-5)^2 = (2)^2 + (2)^2 = 4+4 = 8$

$AC^2 = (9-4)^2 + (7-2)^2 = (5)^2 + (5)^2 = 25+25 = 50$

So, $AB = \sqrt{18} = \sqrt{9 \times 2} = 3\sqrt{2}$; $BC = \sqrt{8} = \sqrt{4 \times 2} = 2\sqrt{2}$;

$CA = \sqrt{50} = \sqrt{25 \times 2} = 5\sqrt{2}$

This gives $AB + BC = 3\sqrt{2} + 2\sqrt{2} = 5\sqrt{2}$

Hence the points *A*, *B* and *C* are collinear.

Example 7

Determine whether the points are vertices of a right triangle

A (–3, –4), *B* (2, 6) and *C* (–6, 10)

Solution Using the distance formula $d = \sqrt{(\boldsymbol{x_2} - \boldsymbol{x_1})^2 + (\boldsymbol{y_2} - \boldsymbol{y_1})^2}$

$AB^2 = (2- -3)^2 + (6 - -4)^2 = (5)^2 + (10)^2 = 25+100 = 125$

$BC^2 = (-6-2)^2 + (10-6)^2 = (-8)^2 + (4)^2 = 64+16 = 80$

$AC^2 = (-6 - -3)^2 + (10- -4)^2 = (-3)^2 + (14)^2 = 9+196 = 205$

$\therefore\ AB^2 + BC2 = 125 + 80 = 205 = CA^2$

Hence *ABC* is a right angled triangle since ***the square of one side is equal to sum of the squares of the other two sides.***

Example 8

Show that the points (*a*, *a*), (- *a*, - *a*) and(- $a\sqrt{3}$, $a\sqrt{3}$) form an equilateral triangle.

Solution

Let the points be represented by A= (*a*, *a*), B= (- *a*, - *a*) and C= (- $a\sqrt{3}$, $\sqrt{3}$) Using the distance formula

$d = \sqrt{(\boldsymbol{x_2} - \boldsymbol{x_1})^2 + (\boldsymbol{y_2} - \boldsymbol{y_1})^2}$

$AB = \sqrt{(-a - a)^2 + (-a - a)^2}$

$=\sqrt{(-2a)^2 + (-2a)^2}$

$=\sqrt{4\,a^2 + 4\,a^2}$

$= \sqrt{8a^2} = 2\sqrt{2a}$

$BC=\sqrt{\left(-a\sqrt{3} - -a\right)^2 + \left(a\sqrt{3 -}- a\right)^2}$

$=\sqrt{\left(-a\sqrt{3} + a\right)^2 + \left(a\sqrt{3 + a}\right)^2}$

$=\sqrt{3a^2 + a^2 - 2a^2\sqrt{3} + 3a^2 + a^2 + 2a^2\sqrt{3}}$

$\sqrt{8a^2} = 2\sqrt{2a}$

$AC= \sqrt{\left(a - -a\sqrt{3}\right)^2 + \left(a - a\sqrt{3}\right)^2}$

$=\sqrt{a^2 + 2a^2\sqrt{3} + 3a^2 + a^2 - 2a^2\sqrt{3} + 3a^2}$

$= \sqrt{8a^2} = 2\sqrt{2a}$

∴ AB = BC= AC

Since ***all the sides are equal*** the points form an equilateral triangle

Example 9

Prove that the points (–7, –3), (5, 10), (15, 8) and (3, –5) taken in order are the corners of a parallelogram.

Solution

Let *A*, *B*, *C* and *D* represent the points (-7, -3), (5, 10), (15, 8) and (3, -5) respectively.

Using the distance formula $d = \sqrt{(\boldsymbol{x_2 - x_1})^2 + (\boldsymbol{y_2 - y_1})^2}$

AB^2= (5- -7)^2+ (10 - -3)^2 = (12)^2 + (13)^2 = 144+169 = 313

BC^2= (15-5)^2+ (8-10)^2 = (10)^2 + (-2)^2 = 100+4 = 104

CD^2= (3-15)^2+ (-5- 15)^2 = (-12)^2 + (-13)^2 = 144+196= 313

$DA^2=(3--7)^2+(-5--3)^2=(10)^2+(-2)^2=100+4=104$

So, $AB=CD=\sqrt{313}$

$BC=DA=\sqrt{104}$

i.e,. ***The opposite sides are equal***. Hence *ABCD* is a parallelogram.

Example 10

Show that the following points (3, –2), (3, 2), (–1, 2) and (–1, –2) taken in order are vertices of a square.

Solution

Let the vertices be taken as *A* (3, -2), *B* (3, 2), *C* (-1, 2) and *D* (-1, -2).

$AB^2=(3-3)^2+(2--2)^2=(0)^2+(4)^2=0+16=16$

$BC^2=(3+1)^2+(2-2)^2=(4)^2+(0)^2=16+0=16$

$CD^2=(-1+1)^2+(2+2)^2=(0)^2+(4)^2=0+16=16$

$DA^2=(-1-3)^2+(-2--2)^2=(-4)^2+(0)^2=16+0=16$

$AB=BC=CD=DA=\sqrt{16}=4$. (That is, all the sides are equal.)

$AC^2=(-1-3)^2+(2--2)^2=(-4)^2+(4)^2=16+16=32$

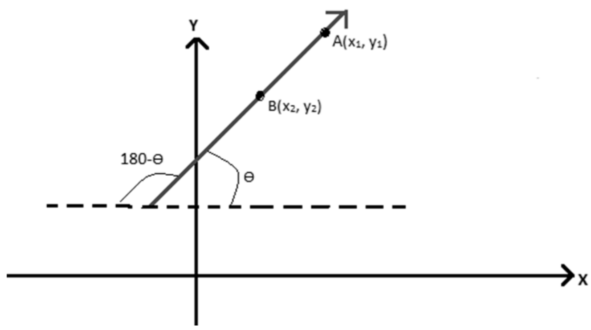
$BD^2=(-1-3)^2+(-2-2)^2=(-4)^2+(-4)^2=16+16=32$

$AC=BD=\sqrt{32}=4\sqrt{2}$ (that is the diagonals are equal)

Hence the points A,B,C and D form a square .

SLOPE OF A STRAIGHT LINE

The measure of steepness and direction of straight line is given by its slope. Slope is usually represented by the letter *m*.



In the given figure, if the angle of inclination of the given line with the x- axis is ∅, then then the slope of the line is represented by tan ∅.

1. **Slope of the line joining two points**

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

2. **Slope of the line which is parallel to x – axis**

$$m = \frac{y - y}{x_2 - x_1}$$

$$m = \frac{0}{x_2 - x_1}$$

3. **Slope of the line which is parallel to y- axis**

$$m = \frac{y_2 - y_1}{x - x}$$

$$m = \frac{y_2 - y_1}{0}$$

4. **Slope of the line joining the origin and any point**

$$m = \frac{y_1}{x_1}$$

5. **Slope of the equation**

$$m = \frac{-a}{b}$$

a= co–efficient of x

b=co-efficient of y

Example 11

Find the slope of the lines joining the points

(i) (-1,3) and (2,5)

(ii) (-2,-1) and (1,3)

Solution

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

(i) Let A= (-1,3) and B= (2,5)

Slope of the line AB = $m = \frac{5-3}{2--1}$

$$= \frac{2}{3}$$

(iii) Let A= (-2,-1) and B= (1,3)

Slope of the line AB = $m = \frac{3--1}{1--2}$

$$= \frac{4}{3}$$

Example 12

Find the slope of the line joining the points

(i) (-3,2) and (4,2)

(ii) (2,5) and (2,3)

(iii) (0,0) and (1,2)

Solution

$$m = \frac{y_2-y_1}{x_2-x_1}$$

(i) Let A= (-3,2) and B=(4,2)

Slope of the line AB = $m = \frac{2-2}{4--3}$

$$m = \frac{0}{7} = 0$$

Note: If *m*= 0, it shows us that the line is parallel to x- axis

(ii) Let A= (2,5) and B=(2,3)

Slope of the line AB = $m = \frac{3-5}{2-2}$

$$= \frac{-2}{0} = 0$$

Note: If *m*= 0, it shows us that the line is parallel to *y*- axis

(iv) Let A= (0,0) and B= (1,2)

$$m = \frac{y_1}{x_1}$$

Slope of the line AB = $m = \frac{2}{1}$

$$= \frac{2}{1} = 2$$

Note : we used $m = \frac{y_1}{x_1}$ **formula, because line passes from the origin**

Example 13

Find the slope of the equation of the lines

(i) **2x+5y-4=0**

(ii) **x-4y=3**

(iii) **y=x+1**

Slope of the equation

$$m = \frac{-a}{b}$$

a= co–efficient of x

b=co-efficient of y

(i) **2x+5y-4=0** **a=2** **b=5**

$$m = \frac{-a}{b}$$

$$m = \frac{-2}{5}$$

(ii) **x-4y=3** **a=1** **b=-4**

$$m = \frac{1}{-4}$$

(iii) **y=x+1**

-x+y=1 **a= -1** **b=1**

$$m = \frac{-(-1)}{1}$$

$$m = \frac{1}{1} = 1$$

Example 14

Show that the points (2,-4) (4,-2) and (7,1) are collinear

Solution

Let A= (2,-4) B= (4,-2) and C= (7, 1)

Slope of AB = $\frac{y_2 - y_1}{x_2 - x_1}$

$$= \frac{-2-(-4)}{4-2}$$

$$= \frac{2}{2} = 1$$

Slope of BC = $\frac{y_2 - y_1}{x_2 - x_1}$

$$= \frac{1-(-2)}{7-4}$$

$$= \frac{3}{3} = 1$$

Slope of AB = slope of BC proved

Therefore A,B and C lie on the same line.

Example 15

Find the value of K if the points (K,3) (-6,4) and (-10,5) are collinear.

Let A= (K,3), B= (-6,4) and C= (-10,5)

When A,B and C are collinear

Slope of AB = slope of BC

$$\frac{4-3}{-6-K} = \frac{5-4}{-10-(-6)}$$

$$\frac{1}{-6-K} = \frac{1}{-4}$$

(cross multiplying)

-6- K= - 4

-K = - 4 + 6

- K=2

K= - 2

Example 16

Show that the line joining the points (-2 , 3) and (4 , 2) is parallel to the line joining the points (3,4) and (- 3 , 5)

Solution

Let A= (-2 , 3) and B= (4 , 2)

C= (3,4) and D= (- 3 , 5)

Slope of AB = $\frac{y_2-y_1}{x_2-x_1} = m_1 = \frac{2-3}{4-(-2)} = \frac{-1}{6} = -\frac{1}{6}$

Slope of CD = $\frac{y_2-y_1}{x_2-x_1} = m_2 = \frac{5-6}{-3-3} = \frac{1}{-6} = -\frac{1}{6}$

$m_1 = m_2$ proved

Hence the line AB is parallel to the line CD

Example 17

If the line joining the points (3,2) and (2,-3) is parallel to the line joining the points (4,3) and (2,k). Find the K value

Solution

Let A= (3,2) and B= (2,-3)

C= (4,3) and D= (2,K)

Slope of AB = Slope of CD

$$\frac{-3-2}{2-3} = \frac{K-3}{2-4}$$

$$\frac{-5}{-1} = \frac{K-3}{-2}$$

(cross multiplying)

$10 = -K + 3$

$10-3=-K$

$-K= 7$

$K=-7$

Example 18

Show that the line joining the points (3,- 4) and (2,1) is perpendicular to the line joining the points (-2 , 2) and (3,3)

Let A= (3,- 4) and B= (2,1)

C= (-2 , 2) and D= (3,3)

Slope of AB = $\frac{y_2 - y_1}{x_2 - x_1} = m_1 = \frac{1-(-4)}{2-3} = \frac{5}{-1} = -5$

. Slope of CD = $\frac{y_2 - y_1}{x_2 - x_1} = m_2 = \frac{3-2}{3-(-2)} = \frac{1}{5}$

$m_1 \times m_2 = -1$

$-5 \times \frac{1}{5} = \frac{-5}{5} = -1$ proved

Hence the line AB is perpendicular to the line CD.

Example 19

If the lines joining the points (-3, 4) and (2,-3) is perpendicular to the line joining the points (3,K) and (2, -3). Find the value of K.

Let A= (-3, 4) and B= (2,-3)

C= (3,K) and D= (2, -3)

Slope of AB = $\frac{y_2 - y_1}{x_2 - x_1} = m_1 = \frac{-3-4}{2-(-3)} = \frac{-7}{5}$

$\frac{-7}{5} \times m_2 = -1$

$m_2 = -1 \times \frac{-5}{7}$

$m_2 = \frac{5}{7}$

. Slope of CD = $\frac{y_2 - y_1}{x_2 - x_1} = m_2 = \frac{5}{7}$

$\frac{5}{7} = \frac{-3-K}{2-3}$

$\frac{5}{7} = \frac{-3-K}{-1}$

(Cross multiplying)

-5= -21-7K

-5+21= -7K

16= -7K

$K = \frac{-16}{7}$

EQUATION OF STRAIGHT LINE

(i) If a line is at a distance *a* and parallel to *x*-axis, then the equation of the line is $\boldsymbol{y = \pm a}$.

(ii) If a line is parallel to *y*-axis at a distance *b* from *y*-axis then its equation is $\boldsymbol{x = \pm b}$

(iii) **Point-slope form :** The equation of a line having slope *m* and passing through the point (x_0, y_0) is given by $\boldsymbol{(y - y_0) = m(x - x_0)}$

(iv) **Two-point-form** : The equation of a line passing through two points (x_1, y_1) *and* (x_2, y_2) is given by

$$\boldsymbol{(y - y_1) = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)}$$

(v) **Slope intercept form** : The equation of the line making an intercept *c* on *y*-axis and having slope *m* is given by

$$\boldsymbol{y = mx + c}$$

Note that the value of *c* will be positive or negative as the intercept is made on the positive or negative side of the *y*-axis, respectively

(vi) **Intercept form** : The equation of the line making intercepts *a* and *b* on *x*- and *y*axis respectively is given by

$$\boldsymbol{\frac{x}{a} + \frac{y}{b} = 1}$$

General equation of a line

Any equation of the form A*x* + B*y* + C = 0, where A and B are simultaneously not zero, is called the general equation of a line.

Example 20

Find the equation to the line with slope $\frac{1}{3}$ and y intercepts 4

Slope – intercept

$\boldsymbol{y = mx + c}$ **m=** $\frac{1}{3}$ c= 4

$$\boldsymbol{y = \frac{1}{3}x + 4}$$

$$\boldsymbol{y} = \frac{x + 12}{3}$$

3y= x+12

-x+3y= 12

-x+3y-12=0

Example 21

Find the equation to the line which passes through (-1,3) and has slope $\frac{1}{3}$

Point – slope form

$(\boldsymbol{y - y_0}) = \boldsymbol{m(x - x_0)}$

$(\boldsymbol{y - 3}) = \frac{1}{3}[\boldsymbol{x} - (-\boldsymbol{1})]$

$(\boldsymbol{y - 3}) = \frac{1}{3}[\boldsymbol{x + 1}]$

(Cross multiplying)

3y-3=x+1

-x+3y=1+3

-x+3y=4 ***or***

-x+3y-4=0

Example 22

Find the equation to the line joining the points (0,-3) and (-4,-5) (0,-3) and (-4,-5)

Two point form

$$(\boldsymbol{y - y_1}) = \frac{\boldsymbol{y_2 - y_1}}{\boldsymbol{x_2 - x_1}}(\boldsymbol{x - x_1})$$

$$(y - -3) = \frac{-5 - -3}{-4 - 0}\ (x - 0)$$

$$(y + 3) = \frac{-2}{-4}\ (x)$$

$$(y + 3) = \frac{1}{2}\ (x)$$

2y+6=x

-x+2y+6=0 or -x+2y=-6

Example 23

Find the equation to the line cutting of intercepts -3 and 4 on x and y axis

Two intercepts form

$\boldsymbol{\frac{x}{a} + \frac{y}{b} = 1}$ **a= - 3 and b= 4**

$$\frac{x}{-3} + \frac{y}{4} = 1$$

$$\frac{-4x+3y}{12} = 1$$

-4x+3y= 12 or

-4x+3y-12=0

Business Application of Analytical Geometry

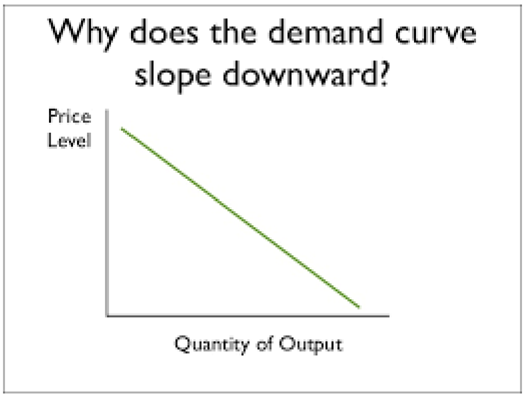
Demand and supply

The law of demand states that there is a negative or inverse relationship between the price and quantity demanded of a commodity over a period of time.

Definition:

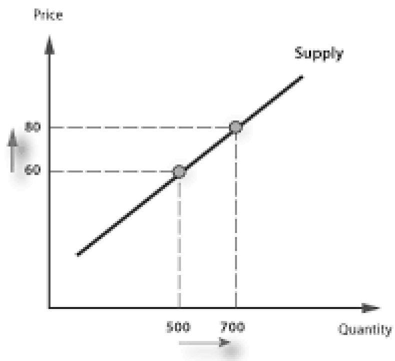
Alfred Marshall stated that "the greater the amount sold, the smaller must be the price at which it is offered, in order that it may find purchasers; or in other words, the amount demanded increases with a fall in price and diminishes with rise in price". According to Ferguson, the law of demand is that the quantity demanded varies inversely with price.

Thus the law of demand states that people will buy more at lower prices and buy less at higher prices, other things remaining the same. By other things remaining the same, we mean the following assumptions.



The demand curve slopes downwards mainly due to the law of diminishing marginal utility. The law of diminishing marginal utility states that an additional unit of a commodity gives a lesser satisfaction. Therefore, the consumer will buy more only at a lower price. ***The demand curve slopes downwards because the marginal utility curve also slopes downwards.***

Supply means the goods offered for sale at a price during a specific period of time. It is the capacity and intention of the producers to produce goods and services for sale at a specific price.



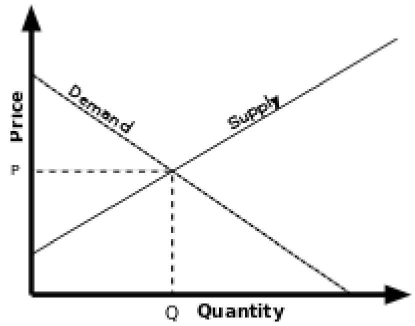
.

The supply of a commodity at a given price may be defined as the amount of it which is actually offered for sale per unit of time at that price.

The law of supply establishes a direct relationship between price and supply. Firms will supply less at lower prices and more at higher prices. "Other things remaining the same, as the price of commodity rises, its supply expands and as the price falls, its supply contracts".

Market Equilibrium

When the supply and demand curves intersect, the market is in equilibrium. This is where the quantity demanded and quantity supplied are equal. The corresponding price is the equilibrium price or market-clearing price, the quantity is the equilibrium quantity.



Example 24

15 tables are sold when the price is Rs 500 and 25 tables are sold when the price is Rs 400. What is equation of the demand curve assuming it to be linear?

Let

X= demand	Y= price
15	500
25	400

Demand curve points are (15, 500) (25, 400)

X_1, Y_1 X_2, Y_2

Equation formula

$$(y - y_1) = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

$$(y - 500) = \frac{400 - 500}{25 - 15}(x - 15)$$

$$(y - 500) = \frac{-100}{10}(x - 15)$$

$$(y - 500) = -10(x - 15)$$

$$(y - 500) = -10x + 150$$

10x+y=150+500

10x+y=650

The demand curve equation is 10x+y=650

Example 25

When the price is Rs 30 , 100 toys of a particular type are available and when the price is Rs 50, 150 toys of the same type are available in the market

Let

X= supply	Y= price
100	30
150	50

Supply curve points are (100, 30) (150, 50)

$X_1, Y_1 \quad X_2, Y_2$

Equation formula

$$(y - y_1) = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

$$(y - 30) = \frac{50 - 30}{150 - 100}(x - 100)$$

$$(y - 30) = \frac{20}{50}(x - 100)$$

$$(y - 30) = \frac{2}{5}(x - 100)$$

Cross multiplying

5(y-30) = 2 (x-100)

5y-150=2x-200

-2x+5y= -200+150

-2x+5y= -50 (or)

2x-5y= 50

The supply curve equation is 2x-5y=50

Example 26

When the price was Rs 500, 50 Radios were available for sale. When the price was Rs 600, 75. Radios were available. What is the supply equation assuming that it is linear. If 100 Radios are made available, what is the expected price per radio?

Let

X= supply	Y= price
50	500
75	600
100	?

Supply curve points are (50, 500) (75, 600)

$X_1,\ Y_1\quad X_2,\ Y_2$

Equation formula

$$(y - y_1) = \frac{y_2 - y_1}{x_2 - x_1}\,(x - x_1)$$

$$(y - 500) = \frac{600 - 500}{75 - 50}\,(x - 50)$$

$$(y - 500) = \frac{100}{25}\,(x - 50)$$

$$(y - 500) = 4\,(x - 50)$$

$$(y - 500) = 4x - 200$$

y = 4x-200+500

y = 4x+300

The supply curve equation is y = 4x+300

When x = 100

y = 4(100)+300

y=700

If 100 Radios are made available the expected price per Radio is Rs 700

Example 27

A firm produces 200 units of the product for a total cost of Rs 730 and 500 units of the product for a total cost of Rs 970.

Assuming the cost curve to be linear, derive the equation of this straight line and use it to estimate the cost of producing 400 units of the product

Let

X= units	Y= cost
200	730
500	970
400	?

Cost curve points are (200, 730) (500, 970)

$X_{1},$ Y_1 $X_{2},$ Y_2

Equation formula

$$(y - y_1) = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

$$(y - 730) = \frac{970 - 730}{500 - 200}(x - 200)$$

$$(y - 730) = \frac{240}{300}(x - 200)$$

$$(y - 730) = \frac{4}{5}(x - 200)$$

$5(y - 730) = 4(x - 200)$

5y-3650= 4x-800

-4x+5y=-800+3650

-4x+5y=2850

The cost curve equation is -4x+5y=2850

When x=400

-4(400)+5y=2850

-1600+5y= 2850

5y=2850+1600

5y=4450

Y=$\frac{4450}{5}$

Y=890

Example 28

The total expenses (y) of a mess are partly constant and partly proportional to the number of members (x) of the mess. The total expenses are Rs 1040 when there are 12 members in the mess and Rs 1600 when there are 20 members. Find (i) the linear relationship between y and x and (ii) the constant expenses and the variable expenses per member.

Let

X= supply	Y= price
12	1040
20	1600

Equation points are (12, 1040) (20, 1600)

$X_{1,}$ Y_1 $X_{2,}$ Y_2

Equation formula

$$(y - y_1) = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

$$(y - 1040) = \frac{1600 - 1040}{20 - 12}(x - 12)$$

$$(y - 1040) = \frac{560}{8}(x - 12)$$

$$(y - 1040) = 70(x - 12)$$

$$(y - 1040) = 70x - 840$$

y=70x-840+1040

y= 70x+200

(i) The linear relationship of y and x is y =70x+200

(ii) Constant expenses is Rs 200

(iii)Variable expenses per member is Rs 70

Exercises

1. Find the point which divides the line segment joining the points (5 , 2) and (7 , 9) internally in the ratio 2 : 7.

Answer **=** $\left(\frac{49}{9}, \frac{32}{9}\right)$

2. Find the point which divides the line segment joining the points (4 , 7) and (1 , 2) externally in the ratio 3 : 2.

Answer =(—5,—8)

3. Find the midpoint of the line segment joining the points (2,-2) and (-1,4)

Answer **=** $\left(\frac{1}{2}, 1\right)$

4. Find the co –ordinates of the centriod of the triangle whose vertices are (2,5) , (5,2) and (6,6)

Answer **=** $\left(\frac{13}{3}, \frac{16}{3}\right)$

5. Find the distance between the points (0, 4) and (6, 8)

Answer **=** $\left(\sqrt{52}\right)$

6. Determine whether the points are vertices of a right angled triangle (0,3),(-2,1),(-1,4)

Answer: $AB^2 = 8, BC^2 = 10, AC^2 = 2$

7. Show that the points (*4*, *-4*), (- *4*, 4) and($4\sqrt{3}$, $4\sqrt{3}$) form an equilateral triangle.

Answer: $AB = 8\sqrt{2},\ BC = 8\sqrt{2},\ AC = 8\sqrt{2}$

8. Prove that the points (–2, –1), (1, 0), (4, 3) and (1, 2) taken in order are the corners of a parallelogram.

Answer: $AB^2 = 10,\ BC^2 = 18,\ CD^2 = 10, AD^2 = 18$

$AB^2 + BC^2 \neq AC^2\ \ BECAUSE\ AC^2 = 52$

9. Show that the following points (3, 2), (5, 4), (3, 6) and (1, 4) taken in order are vertices of a square.

Answer: $AB^2 = 8,\ BC^2 = 8,\ CD^2 = 8, AD^2 = 8$

$AB^2 + BC^2 = AC^2\ \ BECASUSE\ AC^2 = 16$

10. Find the slope of the line joining the points

  i. (3>2)> (-3>1)

  ii. (3>-1)>(-2>0)

  iii. (-2>-1)>(5>7)

Answer: (i) $\frac{-1}{-6}$ ***(ii)*** $\frac{1}{-5}$ ***(iii)*** $\frac{8}{7}$

11. Show that the points (3,-2) (-1,1) and (-5,4) are collinear

Answer: slope of AB= $\frac{3}{-4}$ ***and slope BC =*** $\frac{3}{-4}$

12. Show that the line joining the points (-3,1) and (3 , 4) is parallel to the line joining the points (5,1) and (1 , -1)

Answer: slope of AB= $\frac{1}{2}$ ***and slope BC =*** $\frac{1}{2}$

13. Show that the line joining the points (2,3) and (4,2) is perpendicular to the line joining the points (5 , 3) and (6,5)

Answer: slope of AB= $\frac{-1}{2}$ ***and slope BC =*** $\frac{2}{1}$

Algebraic Expressions

Algebraic expressions are made up of terms like integral or fractional constants, variables and the algebraic operations (addition, subtraction, multiplication and division) between them. However, there can be one or any number of terms forming the expression. Moreover, the terms may be like or unlike terms.

Variables, Constants and Coefficients

Variable

A quantity which can take various numerical values is known as a **variable** (or a **literal**). Variables can be denoted by using the letters *a, b, c, x, y, z,* etc.

Constant

A quantity which has a fixed numerical value is called a **constant**.

For example, 3, -25, $\frac{12}{13}$ and 8.9 are constants.

Term

A term is a constant or a variable or a product of a constant and one or more variables.

Example

$3x^2$, $6x$, -5 are called the terms of the expression $3x^2+6x-5$

In the expression $3x^2+6x-5$ the terms are $3x^2$, $6x$ and -5. The number of terms is 3.

Coefficient

The coefficient of a given variable or factor in a term is another factor whose product with the given variable or factor is the term itself.

If the coefficient is a constant, it is called a constant coefficient or a numerical Coefficient.

Example

In the term $5xy$,

coefficient of xy is 5 (numerical coefficient),

coefficient of $5x$ is y,

coefficient of $5y$ is x.

Depending on the number of terms forming an expression, the algebraic expressions are categorised into the following:

- Monomials
- Binomials
- Trinomials
- Polynomials

Monomial:

An Algebraic expression that contains ***only one term*** is called a monomial.

Example: $\frac{1}{3}$, $3x$, $-3x^2$, $81xyz$, $\frac{5}{11}a^2b$ etc.....

Binomial:

An Algebraic expression that contains ***only two terms*** is called a binomial.

Example: $x+y$, $4a-3b$, $2-3x^2y$, l^2-7m etc.

Trinomial :

An Algebraic expression that contains ***only three terms*** is called a trinomial.

Example: $x+y+z$, $2a-3b+4$, x^2y+y^2z-z etc ....

Polynomial :

An expression containing a finite number of terms with non-zero coefficient is called a polynomial. In other words, it is an expression containing a finite number of terms with the combination of variables, whole number exponents of variables and constants.

Example: $a+b+c+d$, $7xy$, $3abc-10$, $2x+3y-5z$, $3x^5+4x^4-3x^3+72x+5$ etc.

Degree of the Polynomial:

The monomials in the polynomial are called the terms. **The highest power of the terms is the degree of the polynomial**. The coefficient of the highest power of x in a polynomial is called the **leading coefficient** of the polynomial.

Example: $2x^5+4x^4-3x^3+2x^2+72x+5$ is a polynomial in x. Here we have six monomials $2x^5$, $+4x^4$, $-3x^3$, $+2x^2$, $+72x$ and 5 which are called the terms of the polynomial.

Degree of the polynomial is 5.

The leading coefficient of the polynomial is 2

Standard form of the polynomial

A polynomial is in standard form when all the terms are written in order of descending powers of the variables.

Example: $\mathbf{2+9x-9x^2+2x^4-6x^3}$

Now we write the polynomial in the standard form as $\mathbf{2x^4-6x^3-9x^2+9x+2}$.

Like terms and Unlike terms

Terms having the same variable or product of variables with same powers are called **Like terms**. Terms having different variable or product of variables with different powers are called **Unlike terms**.

Example

(i) x, $-5x$, $9x$ are like terms as they have the same variable x

(ii) $4x^2 y$, $7yx^2$ - are like terms as they have the same variable $x^2 y$

Example

(i) $6x$, $6y$ are unlike terms

(ii) $3xy^2$, $5xy$, $8x$, $10y$ - are unlike terms.

Addition and subtraction of expressions

Adding and subtracting like terms

Already we have learnt about like terms and unlike terms.

The basic principle of addition is that we can add only like terms.

To find the sum of two or more like terms, we add the numerical coefficient of the like terms. Similarly, to find the difference between two like terms, we find the difference between the numerical coefficients of the like terms.

There are two methods in finding the sum or difference between the like terms namely,

(i) Horizontal method

(ii) Vertical method

(i) Horizontal method: In this method, we arrange all the terms in a horizontal line and then add or subtract by combining the like terms

Only like or similar terms can be added or subtracted.

Add $2x$ and $5x$.

Solution:

$2x + 5x = (2 + 5) \times x$

$= 7 \times x = 7x$

(ii) Vertical method:

In this method, we should write the like terms vertically and then add or subtract.

Add $4a$ and $7a$.

Solution:

$$\begin{array}{r} 4\,a \\ +\,7\,a \\ \hline 11\,a \\ \hline \end{array}$$

(i) Subtract - $2xy$ from $9xy$.

$$\begin{array}{r} 9xy \\ -2xy \\ \hline 11xy \\ \hline \end{array}$$

(+ change the sign)

Addition and subtraction of polynomial expressions

To find the sum of two or more like terms, we add the numerical coefficient of the like terms. Similarly, to find the difference between two like terms, we find the difference between the numerical coefficients of the like terms.

Problem 1: Add 2x+3 with 3x+5

Answer (2x+3)+(3x+5)

=2x+3+3x+5

=5X+8

Problem 2: Add $5x^2$-6 with $3x^2$-5

Answer $(5x^2-6)+(3x^2-5)$

$= 5x^2-6+3x^2-5$

$=8x^2-11$

Problem 3 subtract 3x+5 from 2x+3

Answer $(2x+3)-(3x+5)$

$=2x+3-3x+5$

$= -x-2$

Problem 4 subtract $3x^2-5$ from $5x^2-6$

Answer $(5x^2-6)-(3x^2-5)$

$=5x^2-6-3x^2-5$

$=2x^2-1$

Polynomial multiplication is a process for multiplying together two or more polynomials. We can perform polynomial multiplication by applying the distributive property to the multiplication of polynomials.

To multiply two polynomials with each other, take the terms of the first polynomial and distribute them over the second polynomial.

$(a+b) (c+d)= a(c+d)+b(c+d)= ac+ad+bc+bd$

Alternatively, distribute the terms of the second polynomial:

$(a+b) (c+d)= (a+b)c+(a+b)d=ac+bc+ad+bd$

Although the terms are in slightly different order, these two results are the same.

Problem 5 Multiply $(3x^2-7x+5) \times (-4x^3)$

$(3x^2-7x+5) \times (-4x^3)= (-4x^3\times3x^2)+(-4x^3\times(-7x)) +(-4x^3\times5)$

$= -12x^5+28x^4+(-20x^3)$

$=-12x^5+28x^4-20x^3$

Problem 6 Multiply $(2x+3) \times (3x-5)$

$(2x+3) \times (3x-5) = [2x\times3x]+[2x (-5)]+[3\times3x]+[3x(-5)]$

$= 6x^2+(-10x)+9x+(-15)$

$=6x^2-10x+9x-15$

$=6x^2-x-15$

Polynomial Division

Polynomial long division is a method/technique by which we can divide a polynomial by another polynomial of the same or a lower degree.

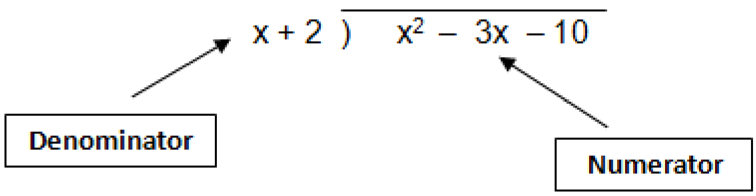
Division of a polynomial (ax^2+bx+c) by another polynomial (dx + e) can be expressed in the form:

$$\frac{ax^2+bx+c}{dx+e}$$

Where a,b,c,d and e are any constant values.

The polynomial on the top is called the "**numerator**" whereas the polynomial on the bottom is termed as "**denominator**". These terms are useful to remember, as we will use them frequently in the coming text. (Note: Remember denominator from down).

While dividing using Long Division method, we write the numerator and the denominator like this:



Problem 7 Divide ($-4x^3$) from ($-12x^5+28x^4-20x^3$)

Answer $= \frac{-12x^5+28x^4-20x^3}{-4x^3} = \frac{-12x^5}{-4x^3} + \frac{28x^4}{-4x^3} - \frac{20x^3}{-4x^3}$

$= 3x^2-7x+5$

Factors of Algebraic Expressions:

If algebraic expressions is expressed as the product of numbers, algebraic variables or algebraic expressions, then each of these numbers and expressions is called the factor of algebraic expressions.

Problem 8 Find the factors of 7x

7x= 7 × x = 7x

=7, x

The factors of 7 and X

Problem 9 Find the factors of 3x-6y

$3x-6y= 3(x-2y)$

$= 3, (x-2y)$

Problem 10 Find the factors of $2x^3+6x^2+4x$

$2x^3+6x^2+4x= 2x(x^2+3x+x)$

$x^2+3x+2= x^2+2x+x+2$

$=x(x+2)+1(x+2)$

$= (x+2) (x+1)$

$2x^3+6x^2+4x= (2x) (x+2) (x+1)$

EQUATIONS

The equation is an expression where two sides of the expression are connected through an equal to sign (=). ***2x+1=9*** is an equation, where ***2x+1*** is the left-hand side (LHS) and 9 is the right-hand side (RHS) of the expression. The equal sign between LHS and RHS indicates that the value of LHS is equal to the RHS of the expression.

10x+63>10, is not an equation. Here, the sign between LHS and RHS of the expression is not an equal sign. Hence, we can say every expression is not an equation.

Some other examples of algebraic equations are, ***8m+5=10n, a+4b=12c+3*** .

NOTE: *An equation is interchangeable i.e. the equation remains same even if LHS and RHS interchange each other.*

Equations ***8m+5=10n*** and ***10n= 8m+5*** are same.

Linear equation

Equations having one variable and the degree of the variable being one is known as a linear equation in one variable. It can be represented by a line parallel to one specific axis. As the number of variables increases, it becomes a linear equation in two variables which becomes more complex. A linear equation in two variables is represented by a line in the cartesian plane varying according to the coefficients of the variable terms.

$$ax + b = 0$$

The standard form of a linear equation in two variables is represented as Where a and b are real numbers, and both a and b are not equal to zero. Every linear equation in one variable has a unique solution.

Both sides of the equation are supposed to be balanced for solving a linear equation. Equality sign denotes that the expressions on either side of the 'equal to' sign are equal. Since the equation is balanced, for solving it certain **mathematical operations** are performed on both sides of the equation in a manner that it does not affect the balance of the equation.

Problem 11 solve 3x+6=0

3x=-6

X=$\frac{-6}{3}$= -2

Problem 12 solve $\frac{x}{2} + 1 = 5$

$\frac{x}{2} = 5 - 1$

$\frac{x}{2} = 4$

$\frac{x}{2} = \frac{4}{1}$ (*Cross multiplying*)

X=8

Problem 13 solve $\frac{x}{3} + \frac{x}{5} = 8$

$\frac{5x+3x}{15}$=8 (*taking L.C.M*)

5x+3x=120

8x=120

X= $\frac{120}{8}$

X= 15

Problem 14 solve $\frac{x+2}{4} = \frac{x-3}{3}$

$\frac{x+2}{4} = \frac{x-3}{3}$ (*Cross multiplying*)

$3(x + 2) = 4(x - 3)$

$3x + 6 = 4x - 12$

$4x - 3x = -12 - 6$

$-x = -18$

$x = 18$

Problem 15 solve $\frac{x}{2} - 2 = \frac{3}{5}x + 4$

Answer $\frac{x}{2} - \frac{3}{5} = 4 + 2$

$\frac{5x-6x}{10} = 6$

$5x - 6x = 60$

$-x = 60$

$x = -60$

Problem 16 solve $\frac{4}{x+2} = \frac{3}{x-3}$

Answer $\frac{4}{x+2} = \frac{3}{x-3}$ *(Cross multiplying)*

$4(x-3) = 3(x+2)$

$4x - 12 = 3x + 6$

$4x - 3x = 6 + 12$

$x = 18$

Problem 17 solve $\frac{x-3}{x-1} + \frac{2x+1}{x-2} = 3$

Answer Multiplying both sides by (x-1) (x-2)

$\frac{(x-3)(x-1)(x-2)}{x-1} + \frac{(2x-1)(x-1)(x-2)}{x-2} = 3(x-1)(x-2)$

$(x^2 - 2x - 3x + 6) + (2x^2 - 2x + x - 1) = 3(x^2 - 2x - x + 2)$

$x^2 - 2x - 3x + 6 + 2x^2 - 2x + x - 1 = 3x^2 - 6x - 3x + 6$

$x^2 + 2x^2 - 3x^2 - 2x - 3x - 2x + x + 6x + 3x = 6 + 1 - 6$

$3x = 1$

$x = \frac{1}{3}$

Simultaneous equations

Simultaneous equations are that equations in which the number of equations must be equal to the number of unknowns. It is used to find the values of unknowns.

Elimination method

In this method, we may combine equations of a system in such a manner as to get rid of one of the unknowns. The elimination of one unknown can be achieved in the following ways.

(i) Multiply or divide the members of the equations by such numbers as to make the coefficients of the unknown to be eliminated numerically equal.

(ii) Then, eliminate by addition if the resulting coefficients have unlike signs and by subtraction if they have like signs.

Substitution method

If two (or more) equations have the same variables and the same solutions then they are simultaneous equations. For example, these equations are simultaneous equations:

$x + y = 3$ and

$2x + 3y = 8$

because both have the **same variables**: 'x' and 'y', and the **same solutions**: $x = 1$, $y = 2$

Substituting $x = 1$ and $y = 2$ into both equations, they BOTH give correct answers:

$1 + 2 = 3$ and

2´1 + 3´2 = 8

Thus: $x = 1$ and $y = 2$ are the solutions to both equations.

'Solving' simultaneous equations means finding the values of 'x' and 'y' that make them true. The following steps will demonstrate how to solve simultaneous equations by the **substitution method**.

We will use the example equations above to demonstrate the procedure...

(1) Isolate one of the variables ('x') on one side of one of the equations:

$x + y = 3$

Isolating 'x':

$x = 3 - y$

(2) Substitute for the isolated variable in the other equation:

$2x + 3y = 8$

Substituting 3 - y for 'x':

$2(3 - y) + 3y = 8$

This equation has only one variable, so we can solve it.

(3) Solve this equation for the other variable, 'y':

$2(3 - y) + 3y = 8$

Expanding the brackets:

$6 - 2y + 3y = 8$

Simplifying:

$6 + y = 8$

Subtracting 6 from both sides:

$y = 2$

(4) Substitute the known value of 'y' into the equation for 'x' derived in step 1:

$x = 3 - y$

Substituting 2 for **'y'**:

$x = 3 - 2$

Therefore:

$x = 1$

Problem 18 solve the equation

x- 2y=1

2x+y=-3

Solution

Method of Elimination

x- 2y=1 ---------------(1)

2x+y=-3---------------(2)

(1) $\times 1$ x - 2y =1

(2) $\times 2$ 4x+2y = -6 (+)

---

(1)+(2) 5x = -5

$x = \frac{-5}{5}$; x= -1

Substituting x= -1 in (1)

x-2y=1

-1-2y=1

-2y=1+1

-2y=2

$$y = \frac{2}{-2} = -1$$

Method of Substitution

x- 2y=1 ----------------(1)

2x+y=-3----------------(2)

x- 2y=1

x=1+2y

Substituting x=1+2y in (2)

2x+y=-3----------------(2)

2(1+2y)+y=-3

2+4y+y=-3

2+5y=-3

5y=-3-2

5y=-5

$$y = \frac{-5}{5} = -1$$

Substituting y= -1 in (1)

x- 2y=1

x- 2(-1)=1

x+2=1

x=1-2

x=-1

Problem 19 Solve the following equation

4x-3y-1=0

2x-5y+3=0

4x-3y=1---------------------(1)

2x-5y=-3--------------------(2)

(1) ×1	4x- 3y = 1	
(2) ×2	4x-10y=-6	(-)
(1)-(2)	7y=7	

$$y = \frac{7}{7} = 1$$

$$y = 1$$

Substituting y= 1 in (1)

4x-3y=1

4x-3(1)=1

4x-3=1

4x=1+3

4x=4

$$x = \frac{4}{4} = 1$$

$$x = 1$$

Method of substitution

4x-3y=1---------------------(1)

2x-5y=-3--------------------(2)

4x-3y=1---------------------(1)

4x-3y=1

4x=1+3y

$x = \frac{1}{4} + \frac{3y}{4}$

Substituting $x = \frac{1}{4} + \frac{3y}{4}$ in (2)

2x-5y=-3--------------------(2)

2($\frac{1}{4} + \frac{3y}{4}$)-5y=-3

$\frac{2}{4} + \frac{6y}{4} - 5y = -3$

$\frac{2 + 6y - 20y}{4} = -3$

$\frac{2-14y}{4} = -3$ (*cross multiplication*)

$2 - 14y = -12$

$-14y = -12 - 2$

$-14y = -14$

$y = \frac{-14}{-14} = 1$

Substituting y= 1 in (1)

4x-3y=1

4x-3(1)=1

4x-3=1

4x=1+3

4x=4

$x = \frac{4}{4} = 1$

$x = 1$

Problem 20 solve $\frac{x}{2} + \frac{y}{3} = 5$

$\frac{x}{4} - \frac{y}{3} = 7$

$$\frac{x}{2} + \frac{y}{3} = 5$$

$$\frac{3x + 2y}{6} = 5$$

$3x + 2y = 30$------------------------------(1)

$$\frac{x}{4} - \frac{y}{3} = 7$$

$$\frac{3x - 4y}{12} = 7$$

$3x - 4y = 84$------------------------------(2)

$3x + 2y = 30$

$3x - 4y = 84$ (-)

6y = -54 (1) - (2)

$$y = \frac{-54}{6} = -9$$

$y = -9$

Substituting y= -9 in (1)

$3x + 2y = 30$

$3x + 2(-9) = 30$

$3x - 18 = 30$

$3x = 30 + 18$

$3x = 48$

$$x = \frac{48}{3} = 16$$

Problem 21

Solve $2x + 3y + z = 11$

$3x + 2y - z = 4$

$x + y - 2z = -3$

$2x + 3y + z = 11$----------------(1)

$3x + 2y - z = 4$------------------(2)

$x + y - 2z = -3$-----------------(3)

$2x + 3y + z = 11$-------------------------(1)
$3x + 2y - z = 4$---------------------------(2) (1) + (2) (z eliminated)

$\mathbf{5x + 5y \quad = 15}$-------------------------(4)

(1) ×2= $4x + 6y + 2z = 22$
(3) ×1= $x + y - 2z = -3$ (1) + (2) (z eliminated)

$\mathbf{5x + 7y \quad = 19}$-------------------(5)

$5x + 5y \quad = 15$ (4)
$5x + 7y \quad = 19$ (5)

$-2y \quad = -4$ (4)-(5)

$y = \frac{-4}{-2} = 2$

$y = 2$

Substituting y= 2 in (4)

$5x + 5y = 15$

$5x + 5(2) \;= 15$

$5x + 10 = 15$

$5x = 15 - 10$

$5x = 5$

$x = \frac{5}{5} = 1$

$x = 1$

Substituting y= 2 and x=1 in (1)

$2x + 3y + z = 11$

$2(1) + 3(2) + z = 11$

$2 + 6 + z = 11$

$8 + z = 11$

$z = 11 - 8$

$z = 3$

Quadratic equation

In math, we define a **quadratic equation** as an **equation** of degree 2, meaning that the highest exponent of this function is 2. The standard form of a**quadratic** is $y = ax^2 + bx + c$, where a, b, and c are numbers and a cannot be 0. Examples of **quadratic equations** include all of these: $y = x^2 + 3x + 1$.

A Quardratic equation is one in which the highest index is 2. It may have i) One variable and ii) Two Variable

Example :

$x^2+7x+8=0$ - One variable

$x^2+2y+y^2=12$ - Two variable

Factorization method

Factorization method can be used when the quadratic equation can be factorized into linear factors. Given a product, if any factor is zero, then the whole product is zero. Conversely, if a product is equal to zero, then some factor of that product must be zero, and any factor which contains an unknown may be equal to zero. Thus, in solving a quadratic equation, we find the values of *x* which make each of the factors zero. That is, we may equate each factor to zero and solve for the unknown.

Formula method

Quadratic formula is the solution of the quadratic equation. In this method, the roots of a quadratic equation are evaluated by using this direct formula.

For example, ax2+bx+c=0is a quadratic equation in standard form and the solution of this quadratic equation is given below

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Problem 22: Solve $x^2+5x+6=0$

Factorisation Method

$x^2+5x+6=0$

1 x 6 = 6

Factors of 6;	Sum of factors
1 x 6	1 + 6 = 7
6 x 1	6 + 1 = 7
2 x 3	2 + 3 = 5
3 x 2	3 + 2 = 5

$x^2+2x+3x+6=0$

x(x+2) + 3(x+2) = 0

(x+2) (x+3)

(x+2)=0 (x+3)=0

X= -2 X= -3

Quadratic formula Method

$x^2+5x+6=0$

a=1; b=5; c=6

$x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$

$x = \frac{-5+\sqrt{5^2-4\times1\times6}}{2\times1}$ $x = \frac{-5-\sqrt{5^2-4\times1\times6}}{2\times1}$

$x = \frac{-5+\sqrt{25-24}}{2}$ $x = \frac{-5-\sqrt{25-24}}{2}$

$x = \frac{-5+\sqrt{1}}{2}$ $\qquad$ $x = \frac{-5-\sqrt{1}}{2}$

$x = \frac{-5+1}{2}$ $\qquad$ $x = \frac{-5-1}{2}$

$x = \frac{-4}{2}$ $\qquad$ $x = \frac{-6}{2}$

$x = -2$ $\qquad$ $x = -3$

Problem 23: solve $2x^2+9x+4=0$

2x4=8

Factors of 8	Sum of factors
1 x 8	1 + 8 = 9
8 x 1	8 + 1 = 9
2 x 4	2 + 4 = 6
4 x 2	4 + 2 = 6

$2x^2+8x+x+4=0$

2x(x+4)+1(x+4)=0

(x+4) (2x+1)=0

(x+4) =0 $\qquad$ (2x+1)=0

X=-4 $\qquad$ 2x=-1; $x = \frac{-1}{2}$

Formation of an equation

To form a quadratic equation, let α and β be the two roots. Let us assume that the required equation be $ax^2 + bx + c = 0$ ($a \neq 0$). According to the problem, roots of this equation are α and β.

Therefore

$(\alpha + \beta) = \frac{-b}{a}$

$\alpha\beta = \frac{c}{a}$

then (x- α) and (x- β) are the factors of the equation.

Hence

$X^2 - X(x- \alpha) + \alpha\beta = 0$

X^2–X(sum of the roots) + product of the roots = 0

Problem 24

From the following equation find

1. The sum of the two roots
2. The product of the two roots

   x-5x+2=0

   a=1; b=-5; c=2

   Sum of the two roots = $\frac{-b}{a}$

   $$= \frac{-(-5)}{1} = 5$$

   Product of the two roots = $\frac{c}{a}$

   $$= \frac{2}{1} = 2$$

Problem 25

. From the following equation find

1. The sum of the two roots
2. The product of the two roots

$2x^2+9x+4=0$

a=2; b=9; c=4

Sum of the two roots = $\frac{-b}{a}$

$$= \frac{-9}{2}$$

Product of the two roots = $\frac{c}{a}$

$$= \frac{4}{2} = 2$$

Example 26

The sum of the two roots is -2 and their product is -35. Find the equation

Sum of the two roots = $\frac{-b}{a} = -2$

= $\frac{-b}{1} = -2;$

$-b = -2$;

$b = 2;$

$a = 1$

Product of the two roots = $\frac{c}{a} = -35$

$\frac{c}{1} = -35,$

$c = -35$

The required equation is $X^2+2x-35=0$

Exercises

1. Add $(3x^2+2x-5)$ with $(8x-7)$ ***Answer*= $\boldsymbol{3x^2 + 10x - 12}$**
2. Subtract $(5x^2+2x-3)$ from $(8x^3+4x^2-3x+5)$ ***Answer*= $\boldsymbol{8x^3 - x^2 - x + 2}$**
3. Multiply $(3x-5) \times (2x+7)$ ***Answer*= $\boldsymbol{6x^2 + 11x - 35}$**
4. Divide $(-4x^3)$ from $(-12x^5+28x^4-20x^3)$ ***Answer*= $\boldsymbol{3x^2 - 7x + 5}$**
5. Find the factors of 7x+14y ***Answer: 7, (x+2y)***
6. Find the factors of $3x^2+3x-6$ ***Answer: 3, (x-1), (x+2)***
7. solve 5x-10=0 ***Answer : x=2***
8. solve $\frac{x}{2} + \frac{x}{7} = 9$ ***Answer : x=14***
9. solve $\frac{x+3}{3} = \frac{x-4}{2}$ ***Answer : x=18***
10. solve the equation

    3x- 4y=-5

    4x+5y=45

Answer: x=5 and y =5

11. solve the equation

$3x-y=1$

$x-2y=-3$

Answer: x=1 and y =2

12. Solve $2x + 2y - 3z = 15$

$x + 3y - 2z = 12$

$3x - y + z = 9$

Answer: x=4, y =2 and z= −1

13. Solve $x^2+8x+15=0$ by two methods

Answer: x=−3 or x =−5

14. solve $3x^2+8x+4=0$ by two methods

Answer: $x=\frac{-2}{3}$ ***or x =−2***

Introduction

We are aware of certain operations of addition and multiplication and now we take up certain higher order operations with powers and roots under the respective heads of indices.

We know that the result of a repeated addition can be held by multiplication e.g.

4 + 4 + 4 + 4 + 4 = 5(4) =

20 a + a + a + a + a = 5(a)

= 5a

Now, $4 \times 4 \times 4 \times 4 \times 4 = 4^5$; $a \times a \times a \times a \times a = a^5$.

It may be noticed that in the first case 4 is multiplied 5 times and in the second case 'a' is multiplied 5 times. In all such cases a factor which multiplies is called the "**base**" and the number of times it is multiplied is called the "**power**" or the "**index**". Therefore, "4" and "a" are the bases and "5" is the index for both. Any base raised to the power zero is defined to be 1; i.e. $a^o = 1$. We also define

$\sqrt[r]{a} = a^{\frac{1}{r}}$

If n is a positive integer, and 'a' is a real number, i.e. n ∈ N and a ∈ R (where N is the set of positive integers and R is the set of real numbers), 'a' is used to denote the continued product of n factors each equal to 'a' as shown below:

$a^n = a \times a \times a$ ………….. to n factors.

Here a^n is a power of "a" whose base is "a" and the index or power is "n".

For example, in $3 \times 3 \times 3 \times 3 = 3^4$, 3 is base and 4 is index or power.

Law of indices

Law 1

$$\mathbf{a^m \times a^n = a^{m+n}}$$

when m and n are positive integers; by the above definition, $a^m = a \times a$ ………….. to m factors and $a^n = a \times a$ ………….. to n factors.

∴ $a^m \times a^n$ = ($a \times a$ ………….. to m factors) × ($a \times a$ ……….. to n factors)

= $a \times a$ ………….. to (m + n) factors

= a^{m+n}

Now, we extend this logic to negative integers and fractions. First let us consider this for negative integer that is m will be replaced by –n. By the definition of $a^m \times a^n = a^{m+n}$, we get $a^{-n} \times a^n = a^{-n+n} = a^0 = 1$

For example

$3^4 \times 3^5 = (3 \times 3 \times 3 \times 3) \times (3 \times 3 \times 3 \times 3 \times 3) = 3^{4+5} = 3^9$

Again, $3^{-5} = 1/3^5 = 1/(3 \times 3 \times 3 \times 3 \times 3) = 1/243$

Law 2

$$\mathbf{a^m/a^n = a^{m-n}}$$

when m and n are positive integers and m > n. By definition, $a^m = a \times a$ ………….. to m factors

Therefore $a^m \div a^n = \frac{a^m}{a^n} = \frac{a \times a \ldots\ldots to\ m\ factors}{a \times a \ldots\ldots to\ n\ factors}$

= $a \times a$ ……… to m–n factors

= a^{m-n}

Now we take a numerical value for a and check the validity of this Law

$$2^7 \div 2^4 = \frac{2^7}{2^4} = \frac{2 \times 2 \ldots\ldots to\ 7\ factors}{2 \times 2 \ldots\ldots to\ 4\ factors}$$

= $2 \times 2 \times 2$ ……….. to (7–4) factors.

= $2 \times 2 \times 2$ ……….. to 3 factors

= $2^3 = 8$

Or

$$2^7 \div 2^4 = \frac{2^7}{2^4} = \frac{2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2}{2 \times 2 \times 2 \times 2}$$

$= 2 \times 2 \times 2 = 2^{1+1+1} = 2^3$

$= 8$

Law 3

$$(a^m)^n =. a^{mn}$$

where m and n are positive integers

By definition $(a^m)^n = a^m \times a^m \times a^m$ …….. to n factors

= (a × a …….. to m factors) a × a ×…….. to n factors……..to n times

= a × a …….. to mn factors

$= a^{mn}$

Following above, $(a^m)^n = (a^m)^{p/q}$

(We will keep m as it is and replace n by p/q, where p and q are positive integers)

Now the qth power of $(a^m)^{p/q}$ is $\{(am)^{p/q}\}^q$

$= (a^m)^{(p/q)x\ q}$

$= a^{mp}$

If we take the qth root of the above we obtain

$$(a^{mp})^{\frac{1}{q}} = \sqrt[q]{a^{mp}}$$

Now with the help of a numerical value for a let us verify this law.

$(2^4)^3 = 2^4 \times 2^4 \times 2^4$

$= 2^{4+4+4}$

$= 2^{12} = 4096$

Law 4

$$(ab)^n = a^n . b^n$$

when n can take all of the values.

For example $6^3 = (2 \times 3)^3 = 2 \times 2 \times 2 \times 3 \times 3 \times 3 = 2^3 \times 3^3$

First, we look at n when it is a positive integer. Then by the definition, we have

$(ab)^n$ = ab × ab ……………. to n factors

= (a × a ……..……. to n factors) × (b × b …………. n factors)

= an × bn

When n is a positive fraction, we will replace n by p/q.

Then we will have $(ab)^n = (ab)^{p/q}$

The qth power of $(ab)^{p/q} = \{(ab)^{(p/q)}\}^q = (ab)^p$

Example 1

Simplify the following

(i) $X^5.X^8$

(ii) $2^3.2^5$

(iii) $X^5 \div X^8$

(iv) $3^5 \div 3^4$

(v) $(a^2.b^3.c)^2$

(vi) $(5^3)^2$

Solution

(i) $X^5.X^8 = X^{5+8}=X^{13}$

(ii) $2^3.2^5= 2^{3+5}=2^8$

(iii) $X^5 \div X^8= \frac{1}{X^{8-5}} = \frac{1}{X^3}$

(iv) $3^5 \div 3^4= 3^{5-4}=3^1 = 3$

(v) $(a^2.b^3.c)^2 = (a^2)^2\ (b^3)^2\ (c)^2$

(vi) $(5^3)^2= (5^{3\times2})= 5^6$

Example 2

Find the value with the help of indices rules

(i) $81^2.27$

(ii) $\frac{256^2}{64^3}$

(iii) (27×216)

(iv) $\frac{1296}{81}$

(v) $(8 \times 16 \times 32)$

Solution

(i) $81^2.27 = (3^4)^2 \cdot 3^3$

$= 3^{8+3}$

$= 3^{11}$

(ii) $\frac{256^2}{64^3} = \frac{(4^4)^2}{(4^4)^3}$

$= \frac{4^8}{4^9} = \frac{1}{4^{9-8}} = \frac{1}{4}$

(iii) $(27 \times 216) = (3^3 \times 6^3) = (3 \times 6)^3 = (18)^3 = 5832$

(iv) $\frac{1296}{81} = \frac{6^4}{3^4} = \left(\frac{6}{3}\right)^4 = 2^4 = 16$

(v) $(8 \times 16 \times 32) = 2^3 . 2^4 . 2^5 = 2^{3+4+5} = 2^{12}$

Example 3

Evaluate

(i) $(\mathbf{5^0})^{\mathbf{2}}$

(ii) $(\mathbf{5^2})^{\mathbf{0}}$

(iii) $\mathbf{5^{(2)^0}}$

(iv) $\boldsymbol{X^0}$

(v) $\boldsymbol{2X^0}$

(vi) $(\boldsymbol{2X})^{\mathbf{0}}$

(vii) $\mathbf{100000^0}$

Solution

(i) $(5^0)^2 = 1^2 = 1$

(ii) $(5^2)^0 = 25^0 = 1$

(iii) $5^{(2)^0} = 5^1 = 5$

(iv) $X^0 = 1$

(v) $2X^0 = 2\times1 = 2$

(vi) $(2X)^0 = 2^0\ X^0 = 1\times1 = 1$

(vii) $100000^0 = 1$

Example 4

Simplify

(i) $7^3 . 6^{-3}$

(ii) $4^{-5} . 4^3$

(iii) $(5^2 . 4^2)^{-3}$

(iv) $\frac{4^{-2}}{5^{-3}}$

(v) $\frac{1}{3^{-2}}$

(vi) $\frac{5^{-2}}{5^{-3}}$

(vii) $.\frac{5^{-3}}{5^{-2}}$

Solution

(i) $7^3 . 6^{-3} = \frac{7^3}{6^3} = \left(\frac{7}{6}\right)^3$

(ii) $4^{-5} . 4^3 = \frac{1}{4^5} . \frac{1}{4^3} = \frac{1}{4^{5+3}} = \frac{1}{4^8} = 4^{-8}$

(iii) $(5^2 . 4^2)^{-3} = (5^2)^{-3}\ (4^2)^{-3} = 5^{-6} . 4^{-6}$

$(5 \times 4)^{-6} = 20^{-6} = \frac{1}{20^6}$

(iv) $\frac{4^{-2}}{5^{-3}} = \frac{1}{4^2} \times \frac{5^3}{1} = \frac{5^3}{4^2} = \frac{125}{16}$

(v) $\frac{1}{3^{-2}} = 1\times\frac{3^2}{1} = 3^2$

(vi) $\frac{5^{-2}}{5^{-3}} = 5^{-2-(-3)} = 5^1 = 5$

(vii) $.\frac{5^{-3}}{5^{-2}} = \frac{1}{5^{-2-(-3)}} = \frac{1}{5^1} = \frac{1}{5}$

Example 5

Simplify $\left(\frac{7}{4}\right)^{-2}$**-**$5(4)^{-3}$**+**$(2X^2)^0$

$= \frac{7^{-2}}{4^{-2}} - 5(4)^{-3} + [2^0 \times (X^2)^0]$

$= \frac{4^2}{7^2} - \frac{5}{4^3} + [1 \times 1]$

$= \frac{16}{49} - \frac{5}{64} + 1$

Exercises

1. Simplify the following

   i. $a^4 \times a^3$ ***Answer :*** $\boldsymbol{a^7}$

   ii. $a^{-3} \times a^5$ ***Answer :*** $\boldsymbol{a^2}$

   iii. $\frac{a^{-3}}{a^5}$ ***Answer :*** $\boldsymbol{\frac{1}{a^8}}$

   iv. $\frac{a^5}{a^{-3}}$ ***Answer :*** $\boldsymbol{a^8}$

   v. $\left(\frac{2x}{5y}\right)^{-2}$ ***Answer :*** $\boldsymbol{\frac{25y^2}{4x^2}}$

   vi. $\left(\frac{7x^2}{-3y}\right)^{-2}$ ***Answer :*** $\boldsymbol{\frac{9y^2}{49x^4}}$

   vii. $(4^3)^{-2}$ ***Answer :*** $\boldsymbol{\frac{1}{4^6}}$

2. Find the value with the help of indices rules

   i. $81^2 \times 27^{-1}$ ***Answer :*** $\boldsymbol{3^5 = 243}$

   ii. $\frac{256^2}{64^{-3}}$ ***Answer :*** $\boldsymbol{2^{34}}$

   iii. $27^{-1} \times 216$ ***Answer :*** $\boldsymbol{\frac{216}{27} = 8}$

Introduction

Sir ***ARTHUR CAYLEY*** (1821-1895) of England was the first Mathematician to introduce the term MATRIX in the year 1858. But in the present day applied Mathematics in overwhelmingly large majority of cases it is used, as a notation to represent a large number of simultaneous equations in a compact and convenient manner. Matrix Theory has its applications in Operations Research, Economics and Psychology. Apart from the above, matrices are now indispensible in all branches of Engineering, Physical and Social Sciences, Business Management, Statistics and Modern Control systems.

Definition of a Matrix

A rectangular array of numbers or functions represented by the symbol.

$\begin{bmatrix} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & a_{2n} \\ a_{m1} & a_{m2} & a_{mn} \end{bmatrix}$ is called a matrix.

The numbers or functions a_{ij} of this array are called elements, may be real or complex numbers, whereas m and n are positive integers, which denotes the number of Rows and number of Columns.

For example

A= $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and B= $\begin{bmatrix} X^2 & \sin x \\ \sqrt{x} & \frac{1}{x} \end{bmatrix}$ are the matrices.

Order of a Matrix

A matrix A with m rows and n columns is said to be of the order m by n (m x n). Symbolically

A = $(a_{ij})_{mxn}$ is a matrix of order m x n. The first subscript i in (a_{ij}) ranging from 1 to m identifies the rows and the second subscript j in (aij) ranging from 1 to n identifies the columns.

For example

A= $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ is a Matrix of order 2×3

B= $\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ is a Matrix of order 2×2

C=$\begin{pmatrix} \sin\theta & \cos\theta \\ \cos\theta & \cos\theta \end{pmatrix}$ is a Matrix of order 2×2

D=$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ is a Matrix of order 3×3

Types of Matrices

(i) SQUARE MATRIX

When the number of rows is equal to the number of columns, the matrix is called a Square Matrix.

B= $\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ is a Matrix of order 2×2

D=$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ is a Matrix of order 3×3

(ii) ROW MATRIX

Amatrix having only one row is called Row Matrix For example

A= (2 0 1) is a row matrix of order 1 x 3

B=(1 0) is a row matrix or order 1 x 2

(iii) COLUMN MATRIX

A matrix having only one column is called Column Matrix. For example

$A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ is a column matrix of order 3×1.

$B = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is a column matrix of order 2×1.

(iv) ZERO OR NULL MATRIX

A matrix in which all elements are equal to zero is called Zero or Null Matrix and is denoted by O.

0=$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is a null matrix of order 2$\times$ 2.

$o = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ is a null matrix of order 3×2

(v) DIAGONAL MATRIX

A square Matrix in which all the elements other than main diagonal elements are zero is called a diagonal matrix

For example

A= $\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$ is a diagonal matrix of order 2 and

B=$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix}$ is a diagonal matrix of order 3

(vi) SCALAR MATRIX

A Diagonal Matrix with all diagonal elements equal to K (a scalar) is called a Scalar Matrix.

For example

A=$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ is a scalar matrix of order 3 and the value of scalar K=2

(vii) UNIT MATRIX OR IDENTITY MATRIX

A scalar Matrix having each diagonal element equal to 1 (unity) is called a Unit Matrix and is denoted by I.

$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is a unit matrix of order 2

$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is a unit matrix of order 3

Multiplication of a marix by a scalar

If A = (a_{ij}) is a matrix of any order and if K is a scalar, then the Scalar Multiplication of A by the scalar k is defined as

KA= (Ka_{ij}) for all i, j.

In other words, to multiply a matrix A by a scalar K, multiply every element of A by K.

Negative of a matrix

The negative of a matrix $A = (a_{ij})_{mxn}$ is defined by $- A = (-a_{ij})_{mxn}$ for all i, j and is obtained by changing the sign of every element.

For example

If $A=\begin{bmatrix} 2 & -5 & 7 \\ 0 & 5 & 6 \end{bmatrix}$ then

$$-A = \begin{matrix} -2 & 5 & -7 \\ -0 & -5 & -6 \end{matrix}$$

Equality of matrices

Two matrices are said to equal when

i) They have the same order and

ii) The corresponding elements are equal.

Addition of matrices

Addition of matrices is possible only when they are of same order (i.e., conformal for addition). When two matrices A and B are of same order, then their sum (A+B) is obtained by adding the corresponding elements in both the matrices.

Properties of matrix addition

Let A, B, C be matrices of the same order. The addition of matrices obeys the following

(i) Commutative law**: A + B = B + A**

(ii) Associative law**: A + (B + C)= (A + B) + C**

(iii) Distributive law: **K (A+B) = KA+KB**, where k is scalar.

Subtraction of matrices

Subtraction of matrices is also possible only when they are of same order. Let A and B be the two matrices of the same order. The matrix A – B is obtained by subtracting the elements of B from the corresponding elements of A.

Multiplication of matrices

Multiplication of two matrices is possible only when the number of columns of the first matrix is equal to the number of rows of the second matrix (i.e. conformable for multiplication)

Let A = (a_{ij}) be an m x p matrix,

and let B = (b_{ij}) be an p x n matrix.

Then the product AB is a matrix C = (c_{ij}) of order m x n,

Properties of matrix multiplication

i. Matrix Multiplication is not commutative i.e. for the two matrices A and B, generally **AB   BA.**

ii. The Multiplication of Matrices is associative i.e., **(AB) C = A(BC)**

iii. Matrix Multiplication is distributive with respect to addition. i.e. if, A, B, C are matrices of order mxn, n x k, and n x k respectively, then **A(B+C) = AB + AC**

iv. Let A be a square matrix of order n and I is the unit matrix of same order.

then **AI = A = I A**

v. The product **AB = O** (Null matrix), does not imply that either A = 0 or B = 0 or both are zero.

Transpose of a matrix

Let A = (a_{ij}) be a matrix of order m x n. The transpose of A, denoted by A^T of order n x m is obtained by interchanging rows into columns of A.

For example

If A = $\begin{bmatrix} 1 & 2 & 5 \\ 3 & 4 & 6 \end{bmatrix}_{2\times3}$ then

$A^T = \begin{bmatrix} 1 & 2 & 5 \\ 3 & 4 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 5 & 6 \end{bmatrix}$

Properties of Matrix Transposition

Let A^T and B^T are the transposed Matrices of A and B and ∝ is a scalar. Then

(i) $(A^T)^T = A$

(ii) $(A + B)^T = A^T + B^T$

(iii) $(\propto A)^T = \propto A^T$

(iv) $(AB)^T = B^T A^T$ (A and B are conformable for multiplication)

Example 1

If A= $\begin{bmatrix} 5 & 9 & 6 \\ 6 & 2 & 10 \end{bmatrix}$ and $B = \begin{bmatrix} 6 & 0 & 7 \\ 4 & -8 & -3 \end{bmatrix}$ **find A + B and A-B**

Solution:

$$A + B = \begin{bmatrix} 5+6 & 9+0 & 6+7 \\ 6+4 & 2+(-8) & 10+(-3) \end{bmatrix}$$

$$A + B = \begin{bmatrix} 11 & 9 & 13 \\ 10 & -6 & 7 \end{bmatrix}$$

$$A - B = \begin{bmatrix} 5-6 & 9-0 & 6-7 \\ 6-4 & 2-(-8) & 10-(-3) \end{bmatrix}$$

$$A - B = \begin{bmatrix} 5-6 & 9-0 & 6-7 \\ 6-4 & 2+8 & 10+3 \end{bmatrix}$$

$$A - B = \begin{bmatrix} -1 & 9 & -1 \\ 2 & 10 & 13 \end{bmatrix}$$

Example 2

If A= $\begin{bmatrix} 3 & 6 \\ 9 & 2 \end{bmatrix}$ find (i) 3A (ii) - $\frac{1}{3}$A

Solution:

(i) 3A= 3 $\begin{bmatrix} 3 & 6 \\ 9 & 2 \end{bmatrix}$ = $\begin{bmatrix} 9 & 18 \\ 27 & 6 \end{bmatrix}$

(ii) $-\frac{1}{3}$ A= $-\frac{1}{3}$ $\begin{bmatrix} 3 & 6 \\ 9 & 2 \end{bmatrix}$ = $\begin{bmatrix} -1 & -2 \\ -3 & -\frac{2}{3} \end{bmatrix}$

Example 3

If A=$\begin{bmatrix} 2 & 3 & 5 \\ 4 & 7 & 9 \\ 1 & 6 & 4 \end{bmatrix}$ and B= $\begin{bmatrix} 3 & 1 & 2 \\ 4 & 2 & 5 \\ 6 & -2 & 7 \end{bmatrix}$

Show that 5(A+B) = 5A+5B

Solution:

$A + B = \begin{bmatrix} 5 & 4 & 7 \\ 8 & 9 & 14 \\ 7 & 4 & 11 \end{bmatrix}$ ∴ 5(A+B) = $\begin{bmatrix} 25 & 20 & 35 \\ 40 & 45 & 70 \\ 35 & 20 & 55 \end{bmatrix}$

$5A = \begin{bmatrix} 10 & 15 & 25 \\ 20 & 35 & 45 \\ 5 & 30 & 20 \end{bmatrix}$ and $5B = \begin{bmatrix} 15 & 5 & 10 \\ 20 & 10 & 25 \\ 30 & -10 & 35 \end{bmatrix}$

$\therefore$ $5A+5B = \begin{bmatrix} 25 & 20 & 35 \\ 40 & 45 & 70 \\ 35 & 20 & 55 \end{bmatrix}$ $\therefore$ 5(A+B) = 5A+5B

Example 4

If $A=\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$ and $B=\begin{bmatrix} -1 & -2 & -4 \\ -1 & -2 & -4 \\ 1 & 2 & 4 \end{bmatrix}$ find AB and BA. Also show that AB ≠ BA

Solution:

$$AB=\begin{bmatrix} 1(-1)+ \quad 2(-1)+ \quad 3(1) & 1(-2)+ \quad 2(-2)+ \quad 3(2) & 1(-4)+ \quad 2(-4)+ \quad 3(4) \\ 2(-1)+ \quad 4(-1)+ \quad 6(1) & 2(-2)+ \quad 4(-2)+ \quad 6(2) & 2(-4)+ \quad 4(-4)+ \quad 6(4) \\ 3(-1)+ \quad 6(-1)+ \quad 9(1) & 3(-2)+ \quad 6(-2)+ \quad 9(2) & 3(-4)+ \quad 6(-4)+ \quad 9(4) \end{bmatrix}$$

$$AB = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{3\times 3}$$

Similarly, $BA= \begin{bmatrix} -17 & -34 & -51 \\ -17 & -34 & -51 \\ 17 & 34 & 51 \end{bmatrix}_{3\times 3}$

$\therefore$ AB ≠ BA

Example 5

If $A= \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix}$ **, then compute A2-5A + 3I**

$A^2= A\times A = \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix} \times \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix} = \begin{bmatrix} -5 & 6 \\ -9 & 10 \end{bmatrix}$

$5A= 5 \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 5 & -10 \\ 15 & -20 \end{bmatrix}$

$3I= 3 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$

□ □ $A^2 - 5A+3I= \begin{bmatrix} -5 & 6 \\ -9 & 10 \end{bmatrix} - \begin{bmatrix} 5 & -10 \\ 15 & -20 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$

$= \begin{bmatrix} -10 & 16 \\ -24 & 30 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} -7 & 16 \\ -24 & 33 \end{bmatrix}$

Example 6

Verify that $(AB)^T = B^T A^T$ when

A= $\begin{bmatrix} 1 & -4 & 2 \\ 4 & 0 & 1 \end{bmatrix}_{2\times3}$ and B = $\begin{bmatrix} 2 & -3 \\ 0 & 1 \\ -4 & -2 \end{bmatrix}_{3\times2}$

Solution:

$$AB = \begin{bmatrix} 1 & -4 & 2 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -3 \\ 0 & 1 \\ -4 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 1(2) + & -4(0) + & 2(-4) & 1(-3) + & -4(1) + & 2(-2) \\ 4(2) + & 0(0) + & 1(-4) & 4(-3) + & 0(1) + & 1(-2) \end{bmatrix}$$

$$= \begin{bmatrix} 2 + & 0 + & -8 & -3 + & -4 + & -4 \\ 8 + & 0 + & -4 & -12 + & 0 + & -2 \end{bmatrix}$$

$$= \begin{bmatrix} -6 & -11 \\ 4 & -14 \end{bmatrix}$$

L.H.S $(\mathbf{AB})^T$ = $\begin{bmatrix} -6 & -11 \\ 4 & -14 \end{bmatrix}^T = \begin{bmatrix} -6 & 4 \\ -11 & -14 \end{bmatrix}$

R.H.S= $B^T A^T = \begin{bmatrix} 2 & 0 & -4 \\ -3 & 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ -4 & 0 \\ 2 & 1 \end{bmatrix}$

$$= \begin{bmatrix} -6 & -11 \\ 4 & -14 \end{bmatrix}$$

$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$ is proved.

DETERMINANTS

An important attribute in the study of Matrix Algebra is the concept of **Determinant,** ascribed to a square matrix. Knowledge of **Determinant** theory is indispensable in the study of Matrix Algebra.

The determinant associated with each square matrix A = (a_{ij}) is a **scalar** and denoted by the symbol det.A or ∣A∣. The scalar may be real or complex number, positive, Negative or Zero. A matrix is an array and has **no numerical** value, but a determinant has **numerical value.**

For example

When A = $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$|A| = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and the determinant value is = **ad- bc**

Example 7

Evaluate $\begin{vmatrix} \mathbf{1} & \mathbf{-1} \\ \mathbf{3} & \mathbf{-2} \end{vmatrix}$

Solution:

$|A| = \begin{bmatrix} 1 & -1 \\ 3 & -2 \end{bmatrix}$

= 1× (–2) – 3 ×(–1) = –2 – (–3)

= –2+3 = 1

Example 8

Evaluate $\begin{vmatrix} \mathbf{2} & \mathbf{0} & \mathbf{4} \\ \mathbf{5} & \mathbf{-1} & \mathbf{1} \\ \mathbf{9} & \mathbf{7} & \mathbf{8} \end{vmatrix}$

Solution:

$\begin{vmatrix} 2 & 0 & 4 \\ 5 & -1 & 1 \\ 9 & 7 & 8 \end{vmatrix} = 2 \begin{vmatrix} -1 & 1 \\ 7 & 8 \end{vmatrix} - 0 \begin{vmatrix} 5 & 1 \\ 9 & 8 \end{vmatrix} + 4 \begin{vmatrix} 5 & -1 \\ 9 & 7 \end{vmatrix}$

= 2(-1×8-1×7) – 0 (5× 8 − 9 × 1)+ 4 (5× 7 − (−1) × 9)

=2(-8-7) – 0 (40-9) + 4 (35+9)

=-30-0+176 = 146

Properties of Determinants

i. The value of determinant is unaltered, when its rows and columns are interchanged.

ii. If any two rows (columns) of a determinant are interchanged, then the value of the determinant changes only in sign.

iii. If the determinant has two identical rows (columns), then the value of the determinant is zero.

iv. If all the elements in a row or in a (column) of a determinant are multiplied by a constant k(k, $\neq 0$) then the value of the determinant is multiplied by k.

v. The value of the determinant is unaltered when a constant multiple of the elements of any row (column), is added to the corresponding elements of a different row (column) in a determinant.

vi. If each element of a row (column) of a determinant is expressed as the sum of two or more terms, then the determinant is expressed as the sum of two or more determinants of the same order.

vii. If any two rows or columns of a determinant are proportional, then the value of the determinant is zero.

Singular Matrix

A square matrix A is said to be singular if det. A = 0, otherwise it is a non-singular matrix.

Example 9

Show that $\begin{vmatrix} \mathbf{1} & \mathbf{2} \\ \mathbf{2} & \mathbf{4} \end{vmatrix}$

Solution:

$\begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} = (1\times 4) - (2\times 2)$

$= 4 - 4$

$= 0$

$\therefore$ The matrix is singular

Example 10

Show that $\begin{vmatrix} \mathbf{2} & \mathbf{5} \\ \mathbf{9} & \mathbf{10} \end{vmatrix}$ **is a non-singular matrix**

Solution:

$\begin{vmatrix} 2 & 5 \\ 9 & 10 \end{vmatrix} = (2\times 10) - (5\times 9)$

$= 20 - 45$

$= -25 \neq 0$

$\therefore$ The given matrix is non singular

INVERSE OF A MATRIX

Minors and Cofactors of the elements of a determinant.

The minor of an element *a*ij of a determinant A is denoted by $M_{i\,j}$ and is the determinant obtained from A by deleting the row and the column where $a_{i\,j}$ occurs.

The cofactor of an element a_{ij} with minor M_{ij} is denoted by C_{ij} and is defined as

$$C_{ij} = \begin{cases} M_{i\,j}\ , if\ \ i+j \ is\ even \\ M_{i\,j}\ , if\ \ i+j\ is\ odd \end{cases}$$

Thus, ***Cofactors are signed minors***

In the case of $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ we have

$M_{11} = a_{22}$; $M_{12} = a_{21}$, $M_{21} = a_{12}$, $M_{22} = a_{11}$

Also $C_{11} = a_{22}$, $C_{12} = - a_{21}$, $C_{21} = - a_{12}$, $C_{22} = a_{11}$

In the case of $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ we have

$M_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$ $\qquad$ $C_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$

$M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$ $\qquad$ $C_{12} = - \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$

$M_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$ $\qquad$ $C_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$

$M_{21} = \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}$ $\qquad$ $C_{21} = - \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}$

Adjoint

The transpose of the matrix got by replacing all the elements of a square matrix A by their corresponding cofactors in | A | is called the **Adjoint** of A or **Adjugate** of A and is denoted by Adj A.

Thus, **AdjA =** $\mathbf{A^T{}_c}$

Note

(i) Let **A =** $\begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{c} & \boldsymbol{d} \end{bmatrix}$ then $\mathbf{A_c} = \begin{bmatrix} \boldsymbol{d} & \boldsymbol{-c} \\ \boldsymbol{-b} & \boldsymbol{a} \end{bmatrix}$

$\therefore$ Adj A = A^T_c = $\begin{bmatrix} \boldsymbol{d} & \boldsymbol{-b} \\ \boldsymbol{-c} & \boldsymbol{a} \end{bmatrix}$

Thus the Adjoint of a 2 x 2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ **c**an be written as $\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

(ii) Adj I = I, where I is the unit matrix.

(iii) A(AdjA) = (Adj A) A = | A | I

(iv) Adj (AB) = (Adj B) (Adj A)

(v) If A is a square matrix of order 2, then |AdjA| = |A|

If A is a square matrix of order 3, then |Adj A| = $|A|^2$

Example 11

Write the Adjoint of the matrix A = $\begin{bmatrix} \mathbf{1} & \mathbf{-2} \\ \mathbf{3} & \mathbf{4} \end{bmatrix}$

Solution

Adj A = $\begin{bmatrix} 4 & -3 \\ 2 & 1 \end{bmatrix}$

Example 12

Find the Adjoint of the matrix A = $\begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{2} \\ \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{3} & \mathbf{1} & \mathbf{1} \end{bmatrix}$

Solution

A = $\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 3 & 1 & 1 \end{bmatrix}$ = Adj A = A^T_c

$C_{11} = \begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} = -1$, $C_{12} = -\begin{vmatrix} 1 & 3 \\ 3 & 1 \end{vmatrix} = 8$, $C_{13} = \begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} = -5$

$C_{21} = -\begin{vmatrix} 1 & 2 \\ 1 & 1 \end{vmatrix} = 1$, ` $C_{22} = \begin{vmatrix} 0 & 2 \\ 3 & 1 \end{vmatrix} = -6$ $C_{23} = -\begin{vmatrix} 0 & 1 \\ 3 & 1 \end{vmatrix} = 3$,

$C_{31} = \begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} = -1$, ` $C_{32} = -\begin{vmatrix} 0 & 2 \\ 1 & 3 \end{vmatrix} = 2$ ` $C_{33} = \begin{vmatrix} 0 & 1 \\ 1 & 2 \end{vmatrix} = -1$

$\therefore \mathbf{A_c} = \begin{bmatrix} -1 & 8 & -5 \\ 1 & -6 & 3 \\ -1 & 2 & -1 \end{bmatrix}$

Hence

$$\text{Adj } A = \begin{bmatrix} -1 & 8 & -5 \\ 1 & -6 & 3 \\ -1 & 2 & -1 \end{bmatrix}^T = \begin{bmatrix} -1 & 1 & -1 \\ 8 & -6 & 2 \\ -5 & 3 & -1 \end{bmatrix}.$$

Inverse of a non singular matrix.

The **inverse of a non singular matrix** A is the matrix B such that AB = BA = I. B is then called the inverse of A and denoted by A^{-1}.

Note

i. A non square matrix has no inverse.

ii. The inverse of a square matrix A exists only when $|A| \neq 0$ that is, if A is a singular matrix then A^{-1} does not exist.

iii. If B is the inverse of A then A is the inverse of B. That is $B = A^{-1}$ and $A = B^{-1}$.

iv. $A A^{-1} = I = A^{-1} A$

v. The inverse of a matrix, if it exists, is unique. That is, no matrix can have more than one inverse.

vi. The order of the matrix A-1 will be the same as that of A.

vii. $I^{-1} = I$

viii. $(AB)^{-1} = B^{-1} A^{-1}$, provided the inverses exist.

ix. $A^2 = I$ implies $A^{-1} = A$

x. If AB = C then

(a) $A = CB^{-1}$ (b) $B = A^{-1}C$, provided the inverses exist.

xi. We have seen that A(AdjA) = (AdjA)A = |A| I

$$\therefore \mathbf{A} \quad \frac{\mathbf{1}}{|\boldsymbol{A}|} \mathbf{(Adj\ A)} = \frac{\mathbf{1}}{|\boldsymbol{A}|} \mathbf{(Adj\ A)\ A = I}$$

This suggests that $A^{-1} = \frac{1}{|A|}(\text{Adj } A)$. That is $A^{-1} = \frac{1}{|A|} A^t_c$

xii. $(A^{-1})^{-1} = A$, provided the inverse exists.

Example 13

Find the inverse of A= $\begin{bmatrix} \mathbf{5} & \mathbf{3} \\ \mathbf{4} & \mathbf{2} \end{bmatrix}$ **if it exists.**

Solution

$|A|=$ $\begin{vmatrix} 5 & 3 \\ 4 & 2 \end{vmatrix}$ = (5× 2) − (3 × 4) = 10-12 = −2

∴ **A^{-1} exists.**

Example 14

Show that the inverse of the following do not exist:

(i) A= $\begin{bmatrix} -2 & 6 \\ 3 & -9 \end{bmatrix}$

(ii) A= $\begin{bmatrix} 3 & 1 & -2 \\ 2 & 7 & 3 \\ 6 & 2 & -4 \end{bmatrix}$

Solution:

(i) $|A|=$ $\begin{bmatrix} -2 & 6 \\ 3 & -9 \end{bmatrix}$ = 0 ∴ **A^{-1} does not exist.**

(ii) $|A|=$ $\begin{bmatrix} 3 & 1 & -2 \\ 2 & 7 & 3 \\ 6 & 2 & -4 \end{bmatrix}$ = 0 ∴ **A^{-1} does not exist.**

Example 15

Find the inverse of A= $\begin{bmatrix} \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{3} & \mathbf{2} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{-2} \end{bmatrix}$ **if it exists**

Solution

$|A|=$ $\begin{bmatrix} \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{3} & \mathbf{2} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{-2} \end{bmatrix}$ **= 15 ≠ 0** ∴ **A^{-1} exists.**

We have $A^{-1} = \frac{1}{|A|} A_c^t$

Now the cofactors are

$C_{11} = \begin{vmatrix} 2 & 1 \\ 1 & -2 \end{vmatrix} = -5$, $C_{12} = -\begin{vmatrix} 3 & 1 \\ 1 & -2 \end{vmatrix} = 7$, $C_{13} = \begin{vmatrix} 3 & 2 \\ 1 & 1 \end{vmatrix} = 1$

$C_{21} = -\begin{vmatrix} 3 & 4 \\ 1 & -2 \end{vmatrix} = 10$, ` $C_{22} = \begin{vmatrix} 2 & 4 \\ 1 & -2 \end{vmatrix} = -8$ $C_{23} = -\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} = 1$

$C_{31} = \begin{vmatrix} 3 & 4 \\ 2 & 1 \end{vmatrix} = -5$, ` $C_{32} = -\begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix} = 10$ ` $C_{33} = \begin{vmatrix} 2 & 3 \\ 2 & 2 \end{vmatrix} = -5$

Hence

$$A_c = \begin{bmatrix} -5 & 7 & 1 \\ 10 & -8 & 1 \\ -5 & 10 & -5 \end{bmatrix}$$

$$A^t_c = \begin{bmatrix} -5 & 10 & -5 \\ 7 & -8 & 10 \\ 1 & 1 & -5 \end{bmatrix}$$

$$\therefore A^{-1} = \frac{1}{|A|} A^t_c$$

$$A^{-1} = \frac{1}{15} \begin{bmatrix} -5 & 10 & -5 \\ 7 & -8 & 10 \\ 1 & 1 & -5 \end{bmatrix}$$

- **If the value of the determinant of a square matrix is not equal to zero, then it is a non-singular matrix.**
- **If it is a non-singular matrix,then inverse exists.**
- **$A^{-1} = \frac{1}{|A|}$ adjoint of A**
- **$A^{-1}B = X$**

Example

Solve the following equation by matrix inverse method

x+2y=6

3x+4y=16

Solution

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 16 \end{bmatrix}$$

A × X= B

X= A^{-1}B

$|A|$ = 1× $\mathbf{4-3\times 2=4-6=-2}$

A^{-1} = $\frac{1}{|A|}$ A^{t}_{c}

adj A = $\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}$

A^{-1} = $-\frac{1}{2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}$

A^{-1}B= $-\frac{1}{2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}\begin{bmatrix} 6 \\ 16 \end{bmatrix}$ = $-\frac{1}{2}\begin{bmatrix} 4\times 6 & + & -2\times 16 \\ -3\times 6 & + & 1\times 16 \end{bmatrix}$

= $-\frac{1}{2}\begin{bmatrix} 24-32 \\ -18+16 \end{bmatrix}$

$-\frac{1}{2}\begin{bmatrix} -8 \\ -2 \end{bmatrix}$

X= $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$

$\left(\frac{x}{y}\right) = \left(\frac{4}{1}\right)$

x= 4

y=1

Example

Solve by matrix method

$2x_1 + 3x_2 - x_3 = 9$

$x_1 + x_2 + x_3 = 9$

$3x_1 - x_2 - x_3 = 1$

Solution

$\begin{bmatrix} 2 & 3 & -1 \\ 1 & 1 & 1 \\ 3 & -1 & -1 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} = \begin{bmatrix} 9 \\ 9 \\ 1 \end{bmatrix}$

A × X= B

X= A^{-1}B

$$A^{-1} = \frac{1}{|A|} A^t_c$$

$$|A| = +(2)\begin{vmatrix} 1 & 1 \\ -1 & -1 \end{vmatrix} - (3)\begin{vmatrix} 1 & 1 \\ 3 & -1 \end{vmatrix} + (-1)\begin{vmatrix} 1 & 1 \\ 3 & -1 \end{vmatrix}$$

$= 2[1 \times (-1) - (-1 \times 1)]$-$3[1 \times (-1) - 3 \times 1] - 1[1 \times (-1) - 3 \times 1]$

=2[-1+1]-3[-1-3]-1[-1+3]

=2(0)-3(-4)-1(-4)

=0+12+4

=16

adj A = A^t_c

$$A^t_c = \begin{pmatrix} +\begin{vmatrix} 1 & 1 \\ -1 & -1 \end{vmatrix} - \begin{vmatrix} 1 & 1 \\ 3 & -1 \end{vmatrix} + \begin{vmatrix} 1 & 1 \\ 3 & -1 \end{vmatrix} \\ -\begin{vmatrix} 3 & -1 \\ -1 & -1 \end{vmatrix} + \begin{vmatrix} 2 & -1 \\ 3 & -1 \end{vmatrix} - \begin{vmatrix} 2 & 3 \\ 3 & -1 \end{vmatrix} \\ +\begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix} - \begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix} + \begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} +(-1+1) - (-1-3) + (-1-3) \\ -(-3-1) + (-2+3) - (-2-9) \\ +(3+1) - (2+1) + (2-3) \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 4 & -4 \\ 4 & 1 & 11 \\ 4 & -3 & -1 \end{pmatrix}$$

$$\text{adj } A = A^t_c = \begin{pmatrix} 0 & 4 & 4 \\ 4 & 1 & -3 \\ -4 & 11 & -1 \end{pmatrix}$$

$$A^{-1} = \frac{1}{16}\begin{pmatrix} 0 & 4 & 4 \\ 4 & 1 & -3 \\ -4 & 11 & -1 \end{pmatrix}$$

$$A^{-1} \cdot B = \frac{1}{16}\begin{pmatrix} 0 & 4 & 4 \\ 4 & 1 & -3 \\ -4 & 11 & -1 \end{pmatrix}\begin{pmatrix} 9 \\ 9 \\ -1 \end{pmatrix}$$

$$= \frac{1}{16}\begin{pmatrix} (0\times 9)+(4\times 9)+(4\times -1) \\ (4\times 9)+(1\times 9)+(-3\times -1) \\ (-4\times 9)+(11\times 9)+(-1\times -1) \end{pmatrix}$$

$$= \frac{1}{16}\begin{pmatrix} 0+36-4 \\ 36+9+3 \\ -36+99+1 \end{pmatrix}$$

$$= \frac{1}{16}\begin{pmatrix} 32 \\ 48 \\ 64 \end{pmatrix}$$

$$X= \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

$x_1=2$

$x_2 =3$

$x_3=4$

Exercises

1. **If A =** $\begin{pmatrix} 1 & 3 & 7 \\ 2 & 4 & 8 \end{pmatrix}$ **and B =** $\begin{pmatrix} 10 & 9 & 6 \\ 11 & 8 & 7 \end{pmatrix}$ **find A+B & A-B**

2. **If A =** $\begin{pmatrix} 0 & 2 \\ 4 & 1 \\ 3 & -2 \end{pmatrix}$ **and B =** $\begin{pmatrix} 1 & -1 \\ 3 & 2 \\ 2 & -3 \end{pmatrix}$ **find A+B & A-B**

3. **If A =** $\begin{pmatrix} 3 & 2 & 1 \\ -1 & -3 & -2 \\ 2 & -1 & 0 \end{pmatrix}$ **find 3A,-A,-3A and** $\frac{1}{2}$**A**

4. **If A =** $\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 3 \end{pmatrix}$ **B=** $\begin{pmatrix} 2 & -1 \\ 0 & 1 \\ 1 & -2 \end{pmatrix}$ **and C=** $\begin{pmatrix} 5 & 0 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}$ **Verify that**

   i. **A+B=B+A**

   ii. **A+(B+C)=(A+B)+C**

   iii. **4(A+B)=4A+AB**

5. If A = $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 5 & 7 \\ 0 & 1 & 4 \end{pmatrix}$ and B = $\begin{pmatrix} 4 & 7 \\ 2 & -9 \\ 1 & 0 \end{pmatrix}$ find AB

6. If A = $\begin{pmatrix} 5 & -3 & 0 \\ 3 & 2 & 1 \\ 3 & 8 & 7 \end{pmatrix}$ and B = $\begin{pmatrix} -8 & 5 & 3 \\ 2 & 7 & 9 \\ 8 & 3 & 0 \end{pmatrix}$ show that AB≠BA

7. If A = $\begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 0 \end{pmatrix}$ and B = $\begin{pmatrix} 1 & 2 \\ 2 & 0 \\ -1 & 1 \end{pmatrix}$ show that AB≠BA

8. If A = $\begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}$ B= $\begin{pmatrix} 1 & 3 & 0 \\ -7 & 6 & -2 \end{pmatrix}$ and C= $\begin{pmatrix} 0 & 1 & 6 \\ 2 & -3 & 5 \end{pmatrix}$ Verify that A(B+C)=AB+AC

9. If A = $\begin{pmatrix} 1 & 0 & -2 \\ 2 & 2 & 4 \\ 0 & 0 & 2 \end{pmatrix}$ show that A^2 -3A+2I = 0

10. If A = $\begin{pmatrix} 1 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix}$ show that $(A^t)^t = A$

11. If A = $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ and B = $\begin{pmatrix} 2 & 2 & 3 \\ 1 & 8 & 6 \end{pmatrix}$ show that

    i. $(A + B)^t = A^t + B^t$

    ii. $(A - B)^t = A^t - B^t$

12. If A = $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ verify that $(3A)^t$ = 3 A^t

13. If A = $\begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}$ find $|A|$

14. Evaluate $\begin{vmatrix} 1 & 1 & 1 \\ 2 & 5 & 7 \\ 2 & 1 & -1 \end{vmatrix}$

15. Find the adjoint of the matrix A= $\begin{pmatrix} 2 & 5 & -2 \\ 4 & -1 & 3 \\ 1 & 3 & -2 \end{pmatrix}$

16. Find the inverse of matrix A$\begin{pmatrix} 2 & 3 \\ 5 & -4 \end{pmatrix}$

17. If A = $\begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix}$ find A^{-1}

18. If A = $\begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 3 \\ 1 & 2 & 1 \end{pmatrix}$ **show that A.** $A^{-1} = A^{-1}.A = I$

19. Solve the following equation by matrix inversion technique

a. *x+2y=6*

b. *3x+4y=16*

20. Solve by matrix method

$2x_1+3x_2-x_3=9$

$x_1+x_2+x_3=9$

$3x_1-x_2-x_3=1$

Answers

1. $A+B=\begin{pmatrix} 11 & 12 & 13 \\ 13 & 12 & 15 \end{pmatrix}$

$A-B=\begin{pmatrix} -9 & -6 & 1 \\ -9 & -4 & 1 \end{pmatrix}$

2. $A+B=\begin{pmatrix} 1 & 1 \\ 7 & 3 \\ 5 & -5 \end{pmatrix}$

$A-B=\begin{pmatrix} -1 & 3 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$

3. $3A=\begin{pmatrix} 9 & 6 & 3 \\ -3 & -9 & -6 \\ 6 & -3 & 0 \end{pmatrix}$

$-3A=\begin{pmatrix} -9 & -6 & -3 \\ 3 & 9 & 6 \\ -6 & 3 & 0 \end{pmatrix}$

$-A=\begin{pmatrix} -3 & -2 & -1 \\ 1 & 3 & 2 \\ -2 & 1 & 0 \end{pmatrix}$

$\frac{1}{2}A=\begin{pmatrix} \frac{3}{2} & 1 & \frac{1}{2} \\ -\frac{1}{2} & -\frac{3}{2} & -1 \\ 1 & -\frac{1}{2} & 0 \end{pmatrix}$

4.

i. **A+B=B+A** **IT *CAN BE PROVED*** $\begin{pmatrix} 3 & 1 \\ 3 & 5 \\ 6 & 1 \end{pmatrix}$

ii. **A+(B+C)=(A+B)+C** **IT *CAN BE PROVED*** $\begin{pmatrix} 8 & 1 \\ 4 & 4 \\ 6 & 2 \end{pmatrix}$

iii. **4(A+B)=4A+AB** **IT *CAN BE PROVED*** $\begin{pmatrix} 12 & 4 \\ 12 & 20 \\ 24 & 4 \end{pmatrix}$

5. **AB =** $\begin{pmatrix} 11 & 25 \\ 29 & 66 \\ 6 & 9 \end{pmatrix}$

6. **AB=** $\begin{pmatrix} -46 & 4 & 42 \\ -12 & 32 & -9 \\ 48 & 92 & -63 \end{pmatrix}$

BA= $\begin{pmatrix} -16 & 58 & 26 \\ 4 & -64 & 56 \\ 49 & -18 & 3 \end{pmatrix}$

7. ***AB***= $\begin{pmatrix} 1 & 4 \\ 4 & 4 \end{pmatrix}$

BA= $\begin{pmatrix} 5 & 3 & 2 \\ 2 & 2 & 4 \\ 1 & 0 & -2 \end{pmatrix}$

8. $\begin{pmatrix} 14 & 10 & 18 \\ -8 & -9 & -15 \end{pmatrix}$

A(B+C)=AB+AC **(*It can be proved)***

9. $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ **A^2 -3A+2I = 0** **(*It can be proved)***

10. A^t $\begin{pmatrix} 1 & 5 \\ 3 & 6 \\ 4 & 7 \end{pmatrix}$ **(*It can be proved)***

11. $(A + B)^t = A^t + B^t$ **(*It can be proved)***

$\begin{pmatrix} 3 & 5 \\ 5 & 13 \\ 7 & 12 \end{pmatrix}$

$(\boldsymbol{A}-\boldsymbol{B})^{t}=\boldsymbol{A}^{t}-\boldsymbol{B}^{t}$ ***(It can be proved)***

$$\begin{pmatrix} -1 & 3 \\ -1 & -3 \\ -1 & 0 \end{pmatrix}$$

12. $(\boldsymbol{3A})^{t}$ **= 3** $\boldsymbol{A}^{t}$ `***(It can be proved)***

$$\begin{pmatrix} 3 & 12 \\ 6 & 15 \\ 9 & 18 \end{pmatrix}$$

13. $|\boldsymbol{A}| = \boldsymbol{-3}$

14. $|\boldsymbol{A}| =$ **-4**

15. Adj A $= \begin{pmatrix} -7 & 4 & 13 \\ 11 & -2 & -14 \\ 13 & -1 & -22 \end{pmatrix}$

16. $\boldsymbol{A}^{-1} = \begin{pmatrix} \frac{4}{23} & \frac{3}{23} \\ \frac{5}{23} & \frac{2}{23} \end{pmatrix}$

17. $\boldsymbol{A}^{-1} = \frac{1}{4}\begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & 1 \\ -1 & 1 & 3 \end{pmatrix}$

18. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ ***(It can be proved)***

19. x= 4 and y= 1

20. $\boldsymbol{x_1 = 2, x_2 = 3, x_3 = 4}$