Bachelor of Commerce

BCOM 302

BUSINESS STATISTICS-1

Directorate of Distance Education
Guru Jambheshwar University of Science & Technology
HISAR-125001

Contents

Subject : **Business Statistics-1**	
Course Code : **BC 302**	Author : **Dr. Pradeep Gupta**
Lesson No. : 1	Vetter: **Prof. B. S. Bodla**
INTRODUCTION TO STATISTICS	

Structure

1.0 Learning Objectives

After going through this lesson, you will be able to

- Explain the concept of statistics
- find the scope of statistics
- find the usefulness of statistics
- find the limitations of statistics
- explain the distrust of statistics

1.1 Introduction

Life in the modern world is inexuricably bound with the notions of number, counting and measurement. One day try to think of a community that cannot count or take measurements and yet is concerned with such acts as selling and buying, carrying on bank transactions, operating locomotives, cars, ships, aircraft and taking part in government. The overriding importance of numerical data in modern life will then be all too apparent. Statistics is being used both is a singular noun and a plural noun.

Statistics, as a plural noun, is used to mean numerical data which arise from a host of uncontrolled, and mostly unknown, causes acting together. It is in this sense that the term statistics is used when our daily newspapers give vital statistics, crime statistics or soccer statistics of Calcutta, or when the Food Minister in the Lok Sabha quotes statistics of sugar exports or those of food grain production.

Used as singular, statistics is a name for the body of scientific methods which are meant for the collection, classification, tabulation, analysis and interpretation of numerical data. But modern literature on the subject does away with any such distinction.

1.2 Concept of Statistics

Statistics is not a new discipline but as old as the human society itself. In the old days statistics was regarded as the 'Science of Statecraft' and was the by-product of the administrative activity of the State. It has been the traditional function of the governments to keep records of population, births, deaths, taxes crop yields and many other types of activities. Counting and measuring these events may generate much kind of numerical data.

The word 'statistics' comes from the Italian word 'statista' (meaning "Statesman") or the German word 'Statistik' each of which means a Political State. It was first used by Professor Gottfried in 1749 to refer to the subject matter as a whole. The science of statistics is said to have originated from two main sources.

(a) *Government Record*: This is the earliest foundation because all cultures with a recorded history had recorded statistics, and the recording, as far as is known, was done by agents of the government for governmental purpose. Since statistical data were collected for governmental purpose, statistics was then described as the 'science of kings' or 'the science of statecraft'.

(b) *Mathematics:* Statistics is said to be a branch of applied mathematics. The present body of statistical methods, particularly those concerned with drawing inferences about population from a sample is based on the mathematical theory of probability.

The following are the two main factors which are responsible for the development or statistics in modern time:

(a) *Increased demand for statistics*: In the present century considerable development has taken place in the field of business and commerce, governmental activities and science. Statistics help in formulating suitable policies, and as such its need is increasingly felt in all these spheres.

(b) *Reduced cost of statistics:* The time and cost of collecting data are very important limiting factors in the use of statistics. However, with the development of electronic machines, such as calculators, computers etc. the cost of analyzing data has considerably gone down. This has led to the increasing use of statistics in solving various problems. Moreover, with the development of statistical theory the cost of collecting and processing data has gone. For example, considerable advance has been made in the sampling techniques which enable us to know the characteristics of the population by studying only a part of it.

1.2.1 Definition of Statistics

The purpose of definition is to lay down precisely the meaning, the scope and the limitations of a subject. There are many definitions of the term 'statistics'. A few definitions are analytically examined below:

(1) Webster defined statistics as "the classified facts representing the conditions of the people in a state especially those facts which can be stated in numbers or in tables of numbers or in any tabular or classified arrangement".

(2) Yule and Kendall defined statistics as "By Statistics we mean quantitative data affected to a marked extent by multiplicity of causes".

(3) Croxton and Cowden have given a very simple and concise definition of statistics. In their view "Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data".

(4) According to Berenson and Levin, "The science of statistics can be viewed as the application of the scientific method in the analysis of numerical data for the purpose of making rational decisions".

(5) Boddington defines statistics as "the science of estimates and probabilities".

(6) According to Lincon L. Chao, "Modern statistics refers to a body of methods and principles that have been developed to handle the collection, description, summarisation and analysis of numerical data. Its primary objective is to assist the researcher in making decisions or generalizations about the nature and characteristics of all the potential observations under consideration of which the collected data form only a small part".

All the above definitions are less comprehensive than the one given by Prof. Horace who defined statistics as follows:

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

(i) *Statistics are aggregate of facts:* Single and isolated figures are not statistics for the simple reason that such figures are unrelated and cannot be compared. To illustrate, if it is stated that the income or Mr. A is Rs. 1, 00,000 per annum, this would not constitute statistics although it is numerical state of fact. Similarly, a single figure relating to production, sale, birth, employment, purchases, accident etc. cannot be regarded statistics although aggregates of such figures would be statistics because of their comparability and relationship as part of common phenomenon.

(ii) *Statistics are affected to a marked extent by multiplicity of causes:* Facts and figures are affected to a considerable extent by a number of forces operating together. For example, statistics of production of rice are affected by the rainfall, quality of soil, seeds, manure, method of cultivation etc.

(iii) *Statistics are numerically expressed:* All statistics are numerical statements of facts i.e. expressed in numbers. Qualitative statements such as 'the population of India is rapidly increasing', or 'the production of wheat is not sufficient' do not constitute statistics. The reason is that such statements are vague and one cannot make at anything from them. On the other hand, the statement 'The estimated population of India at the end of VIIth plan is 803 million' is a statistical statement.

(iv) *Statistics are enumerated or estimated according to a reasonable standard of accuracy:* Facts and figures about any phenomenon can be derived in two ways, viz by actual counting and measurement or by estimate. Estimates cannot be as precise and accurate as actual counts or measurements. The degree of accuracy desired largely depends or measurements. The degree of accuracy desired largely depends upon the nature and object of the enquiry. For example, in measuring heights of persons even 1/10th of a cm is material whereas in measuring distance between two places, say Madras and Calcutta, even fraction of a kilometer can be ignored. However, it is important that reasonable standards of accuracy should be attained; otherwise numbers may be altogether misleading.

(v) *Statistics are collected in a systematic manner:* Before collecting statistics a suitable plan of data collecting should be prepared and the work carried out in a systematic manner. Data

collected in a haphazard manner would very likely lead to fallacious decisions.

(vi) *Statistics are collected for a pre-determined purpose:* The purpose of collecting data must be decided in advance. The purpose should be specific and will define. A general statement of purpose is not enough. For example, if the objective is to collect data on prices, it would not serve any useful purpose unless one knows whether he wants to collect data on wholesale or retail prices and what are the relevant commodities in view.

(vii) *Statistics should be placed in relation to each other:* If numerical facts are to be called statistics, they should be comparable. Statistics data are often compared period-wise or region-wise. For example, the per capita income of India at a particular point of time may be compared with that of earlier years or with the per capita income of other countries, say U.S.A., UK, China etc. Valid comparisons can be made only if the data are homogeneous i.e. relate to the same phenomenon or subject and only likes are compared with likes. It would be meaningless to compare the height of elephants with the height of human beings.

In the absence of the above characteristics, numerical data cannot be called statistics.

1.2.2 Scope of Statistics

(i) *Statistics bring definiteness and precision in conclusions by expressing them numerically.* It is the quality of definiteness which is responsible for the growing universal applications of statistical methods. The conclusions stated numerically are definite and hence more convincing than conclusions stated qualitatively. This fact can be readily understood by a simple example. In an advertisement, statements expressed numerically have greater attention and more appealing than those expressed in a qualitative manner. The caption 'we have sold more. T.Vs this year', is certainly less attractive than 'Record Sale of 15,000 T.V. in 1998 as compared to 10,000 in 1997'. The latter statement emphasizes in a much better manner the growing popularity of the advertises T.Vs.

(ii) *Statistics make data comprehensible to the human mind by simplifying and summarizing it.* Statistics simplifies unwieldy and complex mass of data and presents them in such a manner that they at once become intelligible. The complex data may be reduced to totals, averages, percentage etc. and presented either graphically or diagrammatically. This derives help to understand quickly the significant characteristics of the numerical data, and consequently save

from a lot of mental strain. Single figures in the form of averages and percentages can be grasped more easily than a mass of statistical data comprising thousands of facts. Similarly, diagrams and graphs, because of their greater appeal to the eye and imagination tender valuable assistance in the proper understanding of numerical data. Time and energy of business executives are thus economized, if the statistician supplies them with the results of production, sale and finances in a condensed form.

(iii) Statistics facilitate comparisons in the data. Certain facts, by themselves, may be meaningless unless they are capable of being compared with similar facts at other places or at other periods in times. For example, we estimate the national income of India not essentially for the value of that fact itself, but mainly in order that we may compare the income of today with that of the past and thus draw conclusions as to whether the standard of living of the people is on the increase, decrease or is stationary. It is with the help of statistics that the cost accountant is able to compare the actual accomplishment (in terms of cost). Some of the modes of comparison provided by statistics are: Totals, ratios, averages or measure of central tendencies, graphs & diagrams and coefficients. Statistics thus 'serves as a scale in which facts in various combinations are weighed and valued'.

(iv) Statistics studies and establishes among the variables. Certain statistical measures such as coefficient of correlation, regression etc. establishes relationship between different types of data. For example, it is possible to observe the relationship between income and expenditure, export and forex reserves etc.

(v) Statistics helps in formulating and testing hypothesis. Statistical methods are extremely useful in formulating and testing hypothesis and to develop new theories. For examples the hypothesis that a new drug is effective in checking malaria, will require the use of statistical technique of association of attributes.

(vi) Statistics helps in prediction. Almost all our activities are based on estimates about future and the judicious forecasting of future trends is a prerequisite for efficient implementation of policies. The statistical techniques for extrapolation, time series etc. are highly useful for forecasting future events.

(vii) Statistics helps in the formulation of suitable policies. Statistics help in formulating policies in social, economic and business fields. Various government policies in the field of planning taxation, foreign trade, social security etc. are formulated on the basis of analysis of

statistical data and the inferences drawn from them. For example, vital statistics comprising birth and morality rates help in assessing future growth in population. This information is necessary for designing any scheme of family planning. Similarly, the rate of dearness allowance to be given to the employees is calculated with the help of index numbers.

(viii) *Statistics draws inferences for taking decisions.* Statistical tests are devised to help in drawing valid inference in regard to the nature and characteristics of the universe on the basis of the study of the sample. It can also be the other way when the nature of the sample is judged on the basis of the parameters based on the study of the universe. The validity of such inferences depends on the type of statistical methods employed for the purpose.

(ix) *Statistics endeavors to interpret conditions.* Statistics render useful service by enabling the interpretation of condition, by developing possible causes for the results described. For example, if the production manager discovers that a certain machine is turning out some articles which are not of standard specifications, he will be able to find statistically if this condition is due to some defects in the machine or whether such a condition is normal.

(x) *Statistics measures uncertainty.* Statistical methods help not only in ascertaining the chance of occurrence of an event but also in finding out the total effect of an uncertain event if the consequences of various occurrences are known. Both objective and subjective probability estimates are employed depending upon the nature of the enquiry.

(xi) *Statistics enlarges individual experience.* A proper function of statistics indeed is to enlarge individual experience. Many fields of knowledge would have remained closed to mankind, without the efficient and useful techniques of statistical analysis.

1.2.3 Usefulness of statistics

Statistical methods have become useful tools in the world of affairs. Economy and a high degree of flexibility are the important qualities of statistical methods that render them especially useful to businessmen and scientists.

Statistics and Business: Statistical information is needed from the time the business is launched till the time of its exit. At the time of the floatation of the concern facts are required for the purpose of drawing up the financial plan of the proposed unit. All the factors that are likely to affect judgment on these matters are quantitatively weighed and statistically analyzed before taking the decisions.

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

Correlation and regression analysis help in the estimation of relationships between dependent and one or more independent variables e.g. relationships are established between market demand and per capita income, inputs and outputs etc.

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

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

The use for statistical information in the smooth functioning of an undertaking increases along with its size. The bigger the concern the greater is the need for statistics.

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

Statistics & Economics*:* Statistical data and methods of statistical analysis render valuable assistance in the proper understanding of the economic problems and the formulation of economic policy. Economic problems almost always involve facts that are capable of being expressed numerically, e.g. volume of trade, output of industries - manufacturing, mining and agriculture - wages, prices, bank deposits, clearing house returns etc. These numerical magnitudes are the outcome of a multiplicity of causes and are consequently subject to variations from time to time, or between places or among particular cases. Accordingly, the study of economic problem is specially suited to statistical treatment.

The development of economic theory has also been facilitated by the use of statistics. Statistics is now being used increasingly not only to develop new economic concepts but also to test the old ones. The increasing importance of statistics in the study of economic problem has resulted in a new branch of study called Econometrics.

Statistics and Biology*:* Statistics is being used more and more in biological sciences as an aid to the intelligent planning of experiments, and as a means of assuring the significance of the results of such experiments. Experiments about the growth of animals under different diets and environments, or the crop yields with different seeds, fertilizers and types of soil are frequently designed and analyzed according to statistical principles.

Statistics and physical sciences: Statistics is not much in use in the fields of Astronomy, Geology and Physics. This is due mainly to their relatively high precision of measurements. Statistics has not made any progress in physical sciences beyond the calculation of standard error, and fittings of curves.

Statistics and computers: The development of statistics has been closely related to the evolution of electronic computing machinery. Statistics is a form of data processing, a way of converting data into information useful for decision making. A huge mass of raw data, of related and unrelated nature, derived from internal and external sources of different period of time can be organized and processed into information by computers with accuracy and high speed. The computers can make complex computations, analysis, comparisons and summarizations. Though humans can do the processing, the computer's ability to process huge data is phenomenal, considering its speed, reliability and faithfulness in perfectly following the set of instructions.
The input data in the computer can be processed into a number of different outputs and for a variety of purposes. The system is so organized that managers at different levels and in different activity units are in a position to obtain information in whatever form they want, provided that relevant 'programmes' or instructions have been designed for the purpose. However, the output from a computer is only as good as the data input. 'Garbage In Garbage Out' is an adage familiar to computer users. This warning applies equally to statistical analysis. Statistical decisions based on data are no better than the data used.
As statisticians devise new ways of describing and using data for decisions, computer scientists respond with newer and more efficient ways of performing these operations. Conversely, with the evolution of more powerful computing techniques, people in statistics are encouraged to explore new and more sophisticated methods of statistical analysis.

Statistical Analysis Packages

Statistical Analysis Packages are preprogrammed with all the specialized formulas and built-in procedures a user may need to carry out a range of statistical studies. Statistical programs can:

- Accept data from other sources.
- Add or remove data items, columns or rows.
- Sort, merge and manipulate facts in numerous ways.
- Perform analysis on single and multiple sets of data.
- Convert numeric data into charts and graphs that people can use to grasp relationships, spot patterns and make more informed decisions.
- Print summary values and analysis results.

For a period of at least twenty years, groups of standardized statistical programs assembled as a collection or "package" have been available from various software developers. Recently, there has been a widespread development of statistical packages for use on a microcomputer. Certain packages that were previously available only for mainframe and minicomputers, (such as SAS, SPSS and Minitab) are now available in microcomputer versions, and many new packages (such as STATGRAPHICS, SYSTAT, MYSTAT) have been specifically developed for microcomputer use. The easy and relatively inexpensive access to this type of software has led to its ever-increasing use for business applications.

1.2.4 Limitations of statistics

Though the science of statistics has been profitably applied to an increasingly large number of problems, it has its own limitations and is at times misused by interested people who restrict its scope and utility. According to Newsholme, "It (Statistics) must be regarded as an instrument of research of great value, but having severe limitations which are not possible to overcome and as such they need our careful attention."

The following are some of the important limitations of statistics.

(i) Statistics does not study qualitative phenomenon: Statistics deals with only those subject of inquiry which are capable of being quantitatively measured and numerically expressed. This is an essential condition for the application of statistical methods. Now all subjects cannot be expressed in numbers. Health, poverty, intelligence (to name only a few) is instances of the objects that defy the measuring rod, and hence are not suitable for statistical analysis. The efforts are being made to accord statistical treatment to subjects of this nature also. Health of the people is judged by a study of the death rate, longevity of life and prevalence of any disease or diseases. Similarly intelligence of the students may be compared on the basis of the marks obtained by them in a class test. But these are only indirect methods of approaching the problem and subsidiary to quite a number of other considerations which cannot be

statistically dealth with.

(ii) Statistics does not study individuals: Statistics deals only with aggregates of facts and no importance is attached to individual items. Individual items, taken separately, do not constitute statistical data and are meaningless for any statistical inquiry. For example, the individual figures of agriculture production, industrial output or national income of any country for a particular year are meaningless, unless these figures enable comparison with similar figures for other countries and in the same country these are given for a number of years.

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

(iv) Statistics is only one of the methods of studying a problem: Statistical tools do not provide the best solution under all circumstances. Very often, it is necessary to consider a problem in the light of a country's culture, religion and philosophy, Statistics cannot be of much help in studying such problems. Hence statistical conclusions must be supplemented by other evidences.

(v) Statistics can be misused: The greatest limitation of statistics is that it is liable to be misused. The misuse of statistics may arise because of several reasons. For example, if statistical conclusions are based on incomplete information, one may arrive at fallacious conclusions. Thus the argument that drinking beer is bad for longevity because 99% of the persons who take beer die before the age of 100 years is statistically defective, since we were not told what percentage of persons who do not drink beer die before reaching that age. Statistics are like clay and they can be moulded in any manner so as to establish right or wrong conclusions.

1.3 Distrust of statistics

It is a general belief that "statistics can prove anything." This statement is partly true and false. It is false because mere statistics should not be taken for granted without proper verification. It is true because statistics is often used by unscrupulous people to achieve their personal ends. This results in loss of faith or confidence on statistics or in causing distrust of statistics.

Distrust of statistics literally means lack of trust in statistical data, statistical analysis and the conclusions derived from it. The following reasons account for such views about statistics.

- Facts based on figures are more convincing. But these figures can be manipulated according to one's wishes. This misguides public causing distrust in statistics.
- They can be manipulated in such a manner as to establish foregone conclusions.
- The wrong representation of even correct figures can mislead a reader. Sometimes statistical analyses are misinterpreted causing distrust in statistics. Supposing the mortality rates of patients are more in Indian hospitals. From this one way wrongly conclude that it is safer to treat the patients at home. This type of misinterpretation also causes distrust in statistics.

Statistics are useful tools. One uses them according to his knowledge and experience. Use of statistics makes a statement more convincing. But its misuse causes distrust. So it is necessary that people should be adequately prepared to know the reality or to shift the truth from untruth, good statistics from bad statistics. Thus while working with statistics one should not only avoid outright falsehoods but be alert to detect possible distortion of the truth.

1.4 Check Your Progress

There are some activities to check your progress. Answer the followings:

1. Statistics are affected to a marked extent by multiplicity of …………
2. Statistics helps in the ……………….. of suitable policies.
3. Statistics bring definiteness and ……………in conclusions by expressing them numerically.
4. Distrust of statistics literally means lack of trust in statistical data, statistical analysis and the ………………… derived from it.
5. Statistical data is only…………………………… and not mathematically correct.

1.5 Summary

Statistics is being used both is a singular noun and a plural noun. Statistics, as a plural noun, is used to mean numerical data which arise from a host of uncontrolled, and mostly unknown, causes acting together. Used as singular, statistics is a name for the body of scientific methods which are meant for the collection, classification, tabulation, analysis and interpretation of numerical data. "By statistics we mean aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other". Statistics bring definiteness and precision in conclusions by expressing them numerically. Statistics make data comprehensible to the human mind by simplifying and summarizing it. Statistics facilitate comparisons in the data Statistics studies and establishes among the variables. Statistics helps in formulating, testing hypothesis, prediction and formulation of suitable policies. It is used in every field like business, economics, biology, physical science and computer etc. But there are some limitations of it like Statistics does not study qualitative phenomenon and on the individual basis. It is only approximately and not mathematically correct. It can also be misused Statistics are useful tools. One uses them according to his knowledge and experience. Use of statistics makes a statement more convincing. But its misuse causes distrust. So it is necessary that people should be adequately prepared to know the reality or to shift the truth from untruth, good statistics from bad statistics.

1.6 Keywords

Statistics: It mean aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other.

Distrust of Statistics: It means lack of trust in statistical data, statistical analysis and the conclusions derived from it.

1.7 Self-Assessment Test

Q1. Define statistics. Also discuss the applications of statistics in business decision making.

Q2. Discuss the functions and limitations of statistics.

Q3. "Statistical methods are most dangerous tools in the hands of the expert" Elucidate.

Q4. "Statistics are numerical statement of facts but all facts numerically stated are not statistics" Comment upon the statement and state briefly which numerical statements of facts are not statistics.

Q5. How the computers can be helpful in making statistical decision?

1.8 Answer to Check Your Progress

1. Causes
2. Formulation
3. Precision
4. Conclusions
5. Approximately

1.9 References/Suggested Readings

Gupta, S. P.: Statistical Methods, Sultan chand and sons, New Delhi.

Kumar, S.: Practical Statistics, Sultan chand and sons, New Delhi

Levin, R. and David, S. R.: Statistics for Management, Prentice Hall, New Delhi.

Gupta, C. B.: Introduction to Statistical Methods, Ram Prashad, New Delhi.

Sancheti, D. C. and Kapoor, V. K.: Business Statistics.

Agarwal, B. L.: Basic Statistics, New age International.

Kapur, S. K.: Elements of Practical Statistic, Oxford & IBH Publishers.

Subject : **Business Statistics-1**	
Course Code : **BC 302**	Author : **Ms. Poonam**
Lesson No. : 2	Vetter: **Prof. Suresh K. Mittal**

METHODS OF DATA COLLECTION

Structure:

2.0 Learning Objectives

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

- Know the different methods of data collection
- Understand the methodology of collecting primary data
- Define a questionnaire and its characteristics
- Understand the steps involved in questionnaire designing
- Know designing survey research
- Understand the methodology of collecting secondary data

2.1 Introduction

The facts and figures which can be numerically measured are studied in statistics. Numerical measures of same characteristic are known as observation and collection of observations is termed as data. Data are collected by individual research workers or by organization through sample surveys or experiments, keeping in view the objectives of the study. The data collected may be: Primary Data, Secondary Data. The difference between primary and secondary data in Statistics is that Primary data is collected first hand by a researcher (organization, person, authority, agency or party etc) through experiments, surveys, questionnaires, focus groups, conducting interviews and taking (required) measurements, while the secondary data is readily available (collected by someone else) and is available to the public domain through publications, journals and newspapers.

2.2 Primary Data Collection Methods

Many times due to inadequacy of data or stale information, the need arises for collecting a fresh firsthand information. In marketing research, there are broadly two ways by which primary information can be gathered namely, observation and communication.

Benefits of Primary Data

Benefits of Primary data cannot be neglected. A research can be conducted without secondary data but a research based on only secondary data is least reliable and may have biases because secondary data has already been manipulated by human beings. In statistical surveys it is necessary to get information from primary sources and work on primary data: for example, the statistical records of female population in a country cannot be based on newspaper, magazine and other printed sources. One such source is old and secondly they contain limited information as well as they can be misleading and biased.

Validity: Validity is one of the major concerns in a research. Validity is the quality of a research that makes it trustworthy and scientific. Validity is the use of scientific methods in research which make it logical and acceptable. Using primary data in research can improves the validity of research. Firsthand information obtained from a sample that is representative of the target population will yield data that will be valid for the entire target population.

Authenticity: Authenticity is the genuineness of the research. Authenticity can be at stake if the researcher invests personal biases or uses misleading information in the research. Primary research tools and data can become more authentic if the methods chosen to analyze and interpret data are valid and reasonably suitable for the data type. Primary sources are more authentic because the facts have not been overdone. Primary source can be less authentic if the source hides information or alters facts due to some personal reasons. There are methods that can be employed to ensure factual yielding of data from the source.

Reliability: Reliability is the certainty that the research is enough true to be trusted on. For example, if a research study concludes that junk food consumption does not increase the risk of cancer and heart diseases. This conclusion should have to be drawn from a sample whose size, sampling technique and variability is not questionable. Reliability improves with using primary data. In the similar research mentioned above if the researcher uses experimental method and questionnaires the results will be highly reliable. On the other hand, if he relies on the data available in books and on internet he will collect information that does not represent the real facts.

Limitations of Primary Data Collection

One limitation of primary data collection is that it consumes a lot of time. The researchers will need to make certain preparations in order to handle the different demands of the processes and at the same time, manage time effectively. Besides time consumption, the researchers will collect large volumes of data when they collect primary data. Since they will interact with different people, they will end up with large volumes of data, which they will need to go through when analyzing and evaluating their findings. The primary data also require the greater proportion of workforce to be engaged in the collection of information and analysis, which enhances complexity of operations. There is requirement of large amount of resources to collect primary data. There are several methods of collecting the primary data, which are as follows:

- Observation Method
- Interview Method
- Through Questionnaires
- Through Schedules

Other methods such as warranty cards, distributor audits, pantry audits, consumer panels, using mechanical devices, through projective techniques, deep interviews and content analysis.

2.2.1 Observation Method

In the observation method, only present/current behavior can be studied. Therefore, many researchers feel that this is a great disadvantage. A causal observation could enlighten the researcher to identify the problem. Such as the length of the queue in front of a food chain, price and advertising activity of the competitor etc. Observation is the least expensive mode of data collection.

E*xample:* Suppose a Road Safety Week is observed in a city and the public is made aware of advance precautions while walking on the road. After one week an observer can stand at a street corner and observe the number of people walking on the footpath and those walking on the road during a given period of time. This will tell him whether the campaign on safety is successful or unsuccessful. Sometimes, observation will be the only method available to the researcher.

Types of Observation Methods

There are several methods of observation of which any one or a combination of some of them could be used by the observer. Some of these are:

- Structured or unstructured method
- Disguised or undisguised method
- Direct-indirect observation
- Human-mechanical observation

Structured-Unstructured Observation

Whether the observation should be structured or unstructured depends on the data needed. *Example:* A manager of a hotel wants to know "how many of his customers visit the hotel with their families and how many come as single customers. Here, the observation is structured, since it is clear "what is to be observed". He may instruct his waiter store cord this. This information is required to decide requirements of the chairs and tables and also the ambience.

Disguised-Undisguised Observation

In disguised observation, the respondents do not know that they are being observed. In non- disguised observation, the respondents are well aware that they are being observed. In disguised observation, observers often pose as shoppers. They are known as "mystery shoppers". They are paid by research organizations. The main strength of disguised observation is that it allows for registering the true of the individuals.

Direct-Indirect Observation

In direct observation, the actual behavior or phenomenon of interest is observed. In indirect observation, the results of the consequences of the phenomenon are observed. Suppose, a researcher is interested in knowing about the soft drinks consumption of a student in a hostel room. He may like to observe empty soft drink bottles dropped into the bin. Similarly, the observer may seek the permission of the hotel owner to visit the kitchen or stores. He may carry out a kitchen/stores audit, to find out the consumption of various brands of spice items being used by the hotel. It may be noted that the success of an indirect observation largely depends on "how best the observer is able to identify physical evidence of the problem under study".

Human-Mechanical Observation

Most of the studies in marketing research are based on human observation, wherein trained observers are required to observe and record their observation. In some cases, mechanical devices such as eye cameras are used for observation. One of the major advantages of electrical/ mechanical devices is that their recordings are free from any subjective bias.

Advantages of Observation Method

1. Original data can be collected at the time of occurrence of the event.
2. Observation is done in natural surroundings. Therefore, the facts emerge more clearly, whereas in a questionnaire, experiments have environmental as well as time constraints.
3. Sometimes, the respondents may not like to part with some of the information. Such information can be obtained by the researcher through observation. Observation can also be done on those who cannot articulate.
4. Any bias on the part of the researcher is greatly reduced in the observation method.

Limitations of Observation Method

1. The observer might wait for longer period at the point of observation. And yet the desired event may not take place. Observation is required over a long period of time and hence may not occur.
2. For observation, an extensive training of observers is required.
3. This is an expensive method.
4. External observation provides only superficial indications. To delve beneath the surface is very difficult. Only over behavior can be observed.
5. Two observers may observe the same event, but may draw different inferences.
6. It is very difficult together information on (1) Opinions (2) Intentions.

2.2.2 Interview method

There are different methods of it and which are following:-

Personal Interviews

An interview is called personal when the Interviewer asks the questions face-to-face with the Interviewee. Personal interviews can take place at home, at a shopping mall, on the street, and so on.

Advantages

- The ability to let the Interviewee see, feel and/or taste a product.

- The ability to find the target population. For example, you can find people who have seen a film much more easily outside a theater in which it is playing than by calling phone numbers at random.
- Longer interviews are sometimes tolerated. Particularly with in-home interviews that have been arranged in advance. People may be willing to talk longer face-to-face than to someone on the phone.

Disadvantages

- Personal interviews usually cost more per interview than other methods.
- Change in the characteristics of the population might make sample non-representative.

Telephone Surveys

It is a process of collecting information from sample respondents by calling them over telephone. Surveying by telephone is the most popular interviewing method.

Advantages

- People can usually be contacted faster over the telephone than with other methods.
- You can dial random telephone numbers when you do not have the actual telephone numbers of potential respondents.
- Skilled interviewers can often invite longer or more complete answers than people will give on their own to mail, e-mail surveys.

Disadvantages

- Many telemarketers have given legitimate research a bad name by claiming to be doing research when they start a sales call.
- The growing number of working women often means that no one is at home during the day. This limits calling time to a "window" of about 6-9 p.m. (when you can be sure to interrupt dinner or a favorite TV program).
- You cannot show sample products by phone.

Computer Direct Interviews

These are methods in which the respondents key in(enter)their answers directly in to a computer.

Advantages

- It eliminates data entry and editing costs.
- Answers are more accurate to sensitive questions through a computer than to a person or paper questionnaire.
- Interviewer bias is eliminated. Different interviewers can ask questions in different ways, leading to

different results. The computer asks the questions the same way every time.

Disadvantages

- The interviewees must have access to a computer or it must be provided for them.
- As with mail surveys, computer direct interviews may have serious response rate problems in populations due to literacy levels being low.

E-mail Surveys

Email Questionnaire is a new type of questionnaire system that revolutionizes the way on-line questionnaires are conducted. Unlike other on-line questionnaire systems that need a web server to construct, distribute and manage results, Email Questionnaire is totally email based. It works with the existing email system making on-line questionnaire surveys available to anyone with an Internet connection.

Advantages

- Speed: An email questionnaire can gather several thousand responses within a day or two.
- There are practically no costs involved once the setup has been completed.
- Pictures and sound files can be attached.
- The novelty element of an email survey often stimulates higher response levels than ordinary mail surveys.

Disadvantages

- Researcher must possess or purchase a list of email addresses.
- Some people will respond several times or pass questionnaires along to friends to answer.
- Many people dislike unsolicited email even more than unsolicited regular mail.
- Findings cannot be generalized with email surveys. People who have email are different from those who do not, even when matched on demographic characteristics, such as age and gender.
- Email surveys cannot automatically skip questions or randomize question.

Internet/Intranet (Web Page) Survey

Web surveys are rapidly gaining popularity. They have major speed, cost, and flexibility advantages, but also significant sampling limitations. These limitations restrict the groups that can be studied using this technique.

Advantages

- Web page surveys are extremely fast. A questionnaire posted on a popular Web site can gather several thousand responses within a few hours. Many people who will respond to an email invitation to take a Web survey will do so the first day, and most will do so within a few days.

- There is practically no cost involved once the set up has been completed.
- Pictures can be shown. Some Web survey software can also show video and play sound.
- Web page questionnaires can use complex question skipping logic, randomizations and other features which is not possible with paper questionnaires. These features can assure better data.
- Web page questionnaires can use colors, fonts and other formatting options not possible in most email surveys.
- A significant number of people will give more honest answers to questions about sensitive topics, such as drug use or sex, when giving their answers to a computer, instead of to a person or on paper.
- On an average, people give longer answers to open-ended questions on Web page questionnaires than they do on other kinds of self-administered surveys.

Disadvantages

- Current use of the Internet is far from universal. Internet surveys do not reflect the population as a whole. This is true even if a sample of Internet users is selected to match the general population in terms of age, gender and other demographics.
- People can easily quit in the middle of a questionnaire. They are not as likely to complete along questionnaire on the Web as they would be if talking with a good interviewer.
- Depending on your software, there is often no control over people responding multiple times to bias the results.

Mail Questionnaire

Mail questionnaire is a paper questionnaire, which is sent to selected respondents to fill and post filled questionnaire back to the researcher.

Advantages

1. Easier to reach a larger number of respondents throughout the country.
2. Since the interviewer is not present face to face, the influence of interviewer on the respondent is eliminated.
3. This is the only kind of survey you can do if you have the names and addresses of the target population, but not their telephone numbers.
4. Mail surveys allow the respondent to answer at their leisure, rather than at the often inconvenient moment they are contacted for a phone or personal interview. For this reason, they are not considered

as intrusive as other kinds of interviews.

5. Where the questions asked are such that they cannot be answered immediately, and needs some thinking on the part of the respondent, the respondent can think over leisurely and give the answer.
6. Saves cost (cheaper than interview).
7. No need to train interviewers.
8. Personal and sensitive questions are well answered in this method.
9. The questionnaire can include pictures - something that is not possible over the phone.

Limitations

1. It is not suitable when questions are difficult and complicated. Example, Do you believe in value price relationship?
2. When the researcher is interested in a spontaneous response, this method is unsuitable. Because thinking time allowed to the respondent will influence the answer.
3. In case of a mail questionnaire, it is not possible to verify whether the respondent himself/ herself has filled the questionnaire. If the questionnaire is directed towards the housewife, say, to know her expenditure on kitchen items, she alone is supposed to answer it. Instead, if her husband answers the questionnaire, the answer may not be correct.
4. Any clarification required by the respondent regarding questions is not possible.
5. If the answers are not correct, the researcher cannot probe further.
6. Poor response (30%) - Not all will reply.
7. In populations of lower educational and literacy levels, response rates to mail surveys are often too small to be useful.

2.2.3 Questionnaire

A questionnaire is a research instrument consisting of a series of questions and other prompts for the purpose of gathering information from respondents. The questionnaire was invented by Sir Francis Galton.

Characteristics of Questionnaire

1. It must be simple. The respondents should be able to understand the questions.
2. It must generate replies that can be easily be recorded by the interviewer.
3. It should be specific, so as to allow the interviewer to keep the interview to the point.
4. It should be well arranged, to facilitate analysis and interpretation.
5. It must keep the respondent interested throughout.

Process of Questionnaire Designing

The following are the seven steps involved in designing a questionnaire:

Step 1: Determine What Information is required

The first question to be asked by the market researcher is "what type of information does he need from the survey?" This is valid because if he omits some information on relevant and vital aspects, his research is not likely to be successful. On the other hand, if he collects information which is not relevant, he is wasting his time and money.

At this stage, information required, and the scope of research should be clear. Therefore, the steps to be followed at the planning stage are:

1. Decide on the topic for research.
2. Get additional information on the research issue, from secondary data and exploratory research. The exploratory research will suggest "what are the relevant variables?"
3. Gather what has been the experience with similar study.

Step 2: Different Types of Questionnaire

1. Structured and Non-disguised
2. Structured and Disguised
3. Non-structured and Disguised
4. Non-structured and Non-disguised

Structured and Non-disguised Questionnaire: Here, questions are structured so as to obtain the facts. The interviewer will ask the questions strictly in accordance with the prearranged order. For example, what are the strengths of soap A in comparison with soap B?

(a) Cost is less
(b) Lasts longer
(c) Better fragrance
(d) Produces more lather

1. Structured and non-disguised questionnaire is widely used in market research. Questions are presented with exactly the same wording and same order to all respondents. The reason for standardizing the question is to ensure that all respondents reply the same question. The purpose of the question is clear. The researcher wants the respondent to choose one of the five options given above.

Example:"Subjects attitude towards Cyber laws and the need for government legislation to regulate it".

Certainly, not needed at present   Certainly not needed

I can't say

Very urgently needed

Not urgently needed

2. **Structured and disguised Questionnaire:** This type of questionnaire is least used in marketing research. This type of questionnaire is used to know the peoples' attitude, when a direct undisguised question produces a bias. In this type of questionnaire, what comes out is "what does the respondent know" rather than what he feels. Therefore, the endeavor in this method is to know the respondent's attitude.

Currently, the "Office of Profit" Bill is:

(a) In the Lok Sabha for approval.

(b) Approved by the Lok Sabha and pending in the Rajya Sabha.

(c) Passed by both the Houses, pending the presidential approval.

(d) The bill is being passed by the President.

Depending on which answer the respondent chooses, his knowledge on the subject is classified.

In a disguised type, the respondent is not informed of the purpose of the questionnaire. Here the purpose is to hide "what is expected from the respondent?"

Example: "Tell me your opinion about Mr. Ben's healing effect show conducted at Bangalore?"

"What do you think about the Babri Masjid demolition?"

3. **Non-Structured and Disguised Questionnaire:** The main objective is to conceal the topic of enquiry by using a disguised stimulus. Though the stimulus is standardized by the researcher, the respondent is allowed to answer in an unstructured manner. The assumption made here is that individual's reaction is an indication of respondent's basic perception. Projective techniques are examples of non-structured disguised technique. The techniques involve the use of a vague stimulus, which an individual is asked to expand or describe or build a story, three common types under this category are(a) Word association (b)Sentence completion (c) Storytelling.

4. **Non-structured and Non disguised Questionnaire*:*** Here the purpose of the study is clear, but the responses to the question are open-ended. Example: "How do you feel about the Cyber law currently in practice and its need for further modification"? The initial part of the question is consistent. After presenting the initial question, the interview becomes much unstructured as the interviewer probes more deeply. Subsequent answers by the respondents determine the direction the interviewer takes next. The question asked by the interviewer varies from person to person. This method is called "the depth interview". The major advantage of this method is the freedom permitted to the interviewer. By not

restricting the respondents to a set of replies, the experienced interviewers will be above to get the information from the respondents fairly and accurately.

Step 3: Type of Questions

Open-ended Questions

These are questions where respondents are free to answer in their own words. Example: "What factor do you consider while buying a suit"? If multiple choices are given, it could be color, price, style, brand, etc., but some respondents may mention attributes which may not occur to the researcher. Therefore, open-ended questions are useful in exploratory research, where all possible alternatives are explored. The greatest disadvantage of open-ended questions is that the researcher has to note down the answer of the respondents verbatim. Therefore, there is a likelihood of the researcher failing to record some information. Another problem with open-ended question is that the respondents may not use the same frame of reference.

Example: "What is the most important attribute in a job?"

Ans: Pay

The respondent may have meant "basic pay" but interviewer may think that the respondent is talking about "total pay including dearness allowance and incentive". Since both of them refer to pay, it is impossible to separate two different frames.

Dichotomous Question

These questions have only two answers, 'Yes' or 'no' , 'true' or 'false' 'use' or 'don't use'. Do you use toothpaste? Yes ……….. No…………

There is no third answer. However sometimes, there can be a third answer:

Example: "Do you like to watch movies?"

Ans: Neither like nor dislike.

Dichotomous questions are most convenient and easy to answer. A major disadvantage of dichotomous question is that it limits the respondent's response. This may lead to measurement error.

Close-Ended Questions

There are two basic formats in this type:

- Make one or more choices among the alternatives.
- Rate the alternatives.

Closed-ended questionnaires are easy to answer. It requires less effort on the part of the interviewer. Tabulation and analysis is easier. There are lesser errors, since the same questions are asked to everyone.

The time taken to respond is lesser. We can compare the answer of one respondent to another respondent.

Step 4: Wordings of Questions

Wordings of particular questions could have a large impact on how the respondent interprets them. Even a small shift in the wording could alter the respondent's answer.

Example: "Don't you think that Brazil played poorly in the FIFA cup?" The answer will be 'yes'. Many of them, who do not have any idea about the game, will also most likely say 'yes'. If the question is worded in a slightly different manner, the response will be different.

Example:"Do you think that, Brazil played poorly in the FIFA cup?"This is a straight forward question. The answer could be 'yes', 'no' or 'don't know' depending on the knowledge the respondents have about the game.

Step 5: Sequence and Layout

Some guidelines for sequencing the questionnaire are as follows:

Divide the questionnaire into three parts:

1. Basic information
2. Classification
3. Identification information.

Items such as age, sex, income, education, etc., are questioned in the classification section. The identification part involves body of the questionnaire. Always move from general to specific questions on the topic. This is known as funnel sequence.

Layout: How the questionnaire looks or appears.

Example: Clear instructions, gaps between questions, answers and spaces are part of layout. Two different layouts are shown below:

Layout - 1 How old is your bike?

.....Less than1year........1to2years........2to4years more than 4years.

Layout - 2 how old is your bike?

...... Less than 1 year

.... 1 to 2 years.

........2 to 4 years.

.... More than 4 years.

From the above example, it is clear that layout - 2 is better. This is because likely respondent error due to confusion is minimized.

Therefore, while preparing a questionnaire start with a general question. This is followed by a direct and simple question. This is followed by more focused questions .This will elicit maximum information.

Step 6: Pretesting of Questionnaire

Pretesting of a questionnaire is done to detect any flaws that might be present. For example, the word used by researcher must convey the same meaning to the respondents. Are instructions clear skip questions clear? One of the prime conditions for pretesting is that the sample chosen for pretesting should be similar to the respondents who are ultimately going to participate. Just because a few chosen respondents fill in all the questions going does not mean that the questionnaire is sound.

How Many Questions to be asked? The questionnaire should not be too long as the response will be poor. There is no rule to decide this. However, the researcher should consider that if he were the respondent, how he would react to a lengthy questionnaire. One way of deciding the length of the questionnaire is to calculate the time taken to complete the questionnaire. He can give the questionnaire to a few known people to seek their opinion.

Step 7: Revise and Preparation of Final Questionnaire

Final questionnaire may be prepared after pre testing the questionnaire with the small group of respondents. Questionnaire should be revised for the following:

i. To correct the spellings.
ii. To place the questions in proper order to avoid the contextual bias.
iii. To remove the words which are not familiar to respondents?
iv. To add or remove questions arise in the process of pretest, If any.
v. Top urge the words with double meaning, etc.

2.4 Secondary Data Collection Method

In research, secondary data is collecting and possibly processed by people other than the researcher in question. Common sources of secondary data for social science include censuses, large surveys, and organizational records. In sociology primary data is data you have collected yourself and secondary data is data you have gathered from primary sources to create new research. In terms of historical research, these two terms have different meanings. A primary source is a book or set of archival records .A secondary source is a summary of a book or set of records. Secondary data are statistics that already exist. They have been gathered not for immediate use. This may be described as "those data that have been compiled by some agency other than the user". Secondary data can be classified as:

Internal Secondary Data:-

Internal secondary data is a part of the company's record, for which research is already conducted. Internal data are those that are found within the organization. *Example:* Sales in units, credit outstanding, call reports of sales persons, daily production report, monthly collection report, etc.

External Secondary Data:-

The data collected by the researcher from outside the company. This can be divided into four parts:

1. Census data
2. Individual project report being published
3. Data collected for sale on a commercial basis called syndicated data
4. Miscellaneous data

The following are some of the data that can obtain by census records:

- Census of the wholesale trade
- Census of the retail trade
- Population Census
- Census of manufacturing industries
- Individual project report being published
- Encyclopedia of business information sources
- Product finder
- Thomas registers etc.

Benefits and Limitations of Secondary Data

Benefits

It is far cheaper to collect secondary data than to obtain primary data. For the same level of research budget a thorough examination of secondary sources can yield a great deal of information than can be had through a primary data collection exercise. The time involved in searching secondary sources is much less than that needed to complete primary data collection.

Secondary sources of information can yield more accurate data than that obtained through primary research. This is not always true but where a government or international agency has undertaken a large scale survey, or even a census, this is likely to yield far more accurate results than custom designed and executed surveys when these are based on relatively small sample sizes.

It should not be forgotten that secondary data can play a substantial role in the exploratory phase of the research when the task at hand is to define the research problem and to generate hypotheses. The assembly and analysis of secondary data almost invariably improve the researcher's understanding of the marketing problem, the various lines of inquiry that could or should be followed and the alternative courses of actions which might be pursued.

Secondary sources help define the population. Secondary data can be extremely useful both in defining the population and in structuring the sample to be taken. For instance, government statistics on a country's agriculture will help decide how to stratify a sample and, once sample estimates have been calculated, these can be used to project those estimates to the population.

Limitations

1. *Definition:* The researcher, when making use of secondary data, may misinterpret the definitions used by those responsible for its preparation and drawer roneous conclusions
2. *Measurement error:* When a researcher conducts fieldwork she/he is possibly able to estimate inaccuracies in measurement through the standard deviation and standard error, but these are sometimes not published in secondary sources. The problem is sometimes not so much 'error' but differences in the levels of accuracy required by decision makers.
3. *Sourcebias:* Researchers face the problem of vested interests when they consult secondary sources. Those responsible for their compilation may have reasons for wishing to present a more optimistic or pessimistic set of results for their organization i.e., exaggerated figures or inflated estimates may be stated.
4. *Reliability:* There liability of published statistics may vary over time. Because the systems of collecting data or geographical or administrative boundaries may be changed, or the basis for stratifying a sample may have altered. Other aspects of research methodology that affect the reliability of secondary data is the sample size, response rate, questionnaire design and modes of analysis without any indication of this to the reader of published statistics.
5. *Time scale:* The time period during which secondary data was first compiled may have a substantial effect upon the nature of the data for example: Most censuses take place atten- year intervals, so data from this and other published sources may be out-of-date at the time there searcher wants to make use of the statistics.

2.5 Check Your Progress

1. A major disadvantage of dichotomous question is that it------------the respondent's response.

2. Open-ended questions are useful in .... research, where all possible alternatives are explored.
3. Internal secondary data is a part of the.......... ……………………record.
4. External Secondary Data can be divided into ……………………..parts.
5. Internal data are those that are found………………………….the organization.

2.6 Summary

Primary data may pertain to life style, income, awareness or any other attribute of individuals or groups. There are mainly two ways of collecting primary data namely: (a) Observation (b) By questioning the appropriate sample. Observation method has a limitation i.e., certain attitudes, knowledge, motivation, etc. cannot be measured by this method. For this reason , researcher needs to communicate. Communication method is classified based on whether it is structured or disguised. Questionnaire is easy to administer. This type is most suited for descriptive research. If the researcher wants to do exploratory sturdy, unstructured method is better. In unstructured method questions will have to be framed based on the answer by the respondent. Questionnaire can be administered either in person or online or Mail questionnaire. Each of these methods has advantages and disadvantages. Questions in a questionnaire may be classified into (a) Open question (b) Close ended questions (c) Dichotomous questions, etc. While formulating questions, care has to be taken with respect to question wording, vocabulary; leading, loading and confusing questions should be avoided. Further it is desirable that questions should not be complex, or too long. It is also implied that proper sequencing will enable the respondent to answer the question easily. The researcher must maintain a balanced scale and must use a funnel approach. Pretesting of the questionnaire is preferred before introducing to a large population. Secondary data are statistics that already exists. Secondary data may not be readily used because these data are collected for some other purpose. Secondary data has its own advantages and disadvantages. There are two types of secondary data (1) Internal and (2) External secondary data. Census is the most important among secondary data. Syndicated data is an important form of secondary data. Syndicated data may be classified into (a) Consumer purchase data (b) Retailer and wholesaler data (c) Advertising data. Each has advantages and disadvantages.

2.7 Keywords

1. **Dichotomous question:** These questions have only two answers, like ‘yes or no’.
2. **Disguised observation:** The observation under which the respondents do not know that they are being

observed.

3. **Non-Disguised observation:** The observation in which the respondents are well aware that they are being observed.

4. **External Data:** The data collected by the researcher from outside the company.

5. **Internal Data:** Internal data are those that are found within the organization.

6. **Panel Type Data:** This is one type of syndicated data in which there are consumer panels.

7. **Secondary Data:** Secondary data is collecting and possibly processed by people other than the researcher in question.

8. **Syndicated Data:** Data collected by this method is sold to interested clients on payment.

2.8 Self- Assessment Test

Q1. What is primary data?

Q2. What are the various methods of collecting primary data?

Q3. What is questionnaire? What are its importance and characteristics?

Q4. Explain open ended and close ended questions in a questionnaire.

Q5. What are the advantages and disadvantages of primary data?

Q6. What are the sources of secondary data?

Q7. What are the types of secondary data?

Q8. What are the advantages and disadvantages of secondary data?

2.9 Answers to check Your Progress

1. Limits
2. Exploratory
3. Company's
4. Four
5. Within

2.10 References/ Suggested Readings

Gupta, S. P.: Statistical Methods, Sultan chand and sons, New Delhi.

Kumar, S.: Practical Statistics, Sultan chand and sons, New Delhi

Levin, R. and David, S. R.: Statistics for Management, Prentice Hall, New Delhi.

Gupta, C. B.: Introduction to Statistical Methods, Ram Prashad, New Delhi.

Sancheti, D. C. and Kapoor, V. K.: Business Statistics.

Agarwal, B. L.: Basic Statistics, New age International.

Kapur, S. K.: Elements of Practical Statistic, Oxford & IBH Publishers.

Subject: Business Statistics-1	
Course Code: 302	**Author :** Dr. Pradeep Gupta
Lesson No. : 3	**Vetter:** Dr. B.S. Bodla
MEASURES OF CENTRAL TENDENCY	

Structure

3.0 Learning Objectives

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

- Understand the concept of Arithmetic mean
- Understand the concept of Median
- Understand the concept of Mode
- Understand the concept of Harmonic mean
- Understand the concept of Geometric mean

3.1 Introduction

Central tendency or 'average' value is the powerful tool of analysis of data that represents the entire mass of data. The word 'average' is commonly used in day to day conversation. For example, we often talk of average income, average age of employee, average height, etc. An 'average' thus is a single value which is considered as the most representative or typical value for a given set of data. Such a value is neither the smallest nor the largest value, but is a number whose value is somewhere in the middle of the group. For this reason an average is frequently referred to as a *measure of central tendency of central value.* Measures of central tendency show the tendency of some central value around which the data tends to cluster.

Characteristics of a Good Average:

Since an average is a single value representing a group of values, it is desirable that such a value satisfies the following properties:

(i) It should be easy to understand: Since statistical methods are designed to simplify complexity, it is desirable that an average be such that it can be readily understood, otherwise, its use is bound to be very limited.

(ii) It should be simple to compute: Not only an average should be easy to understand but also it should be simple to compute so that it can be used widely. However, though ease of computation is desirable, it should not be sought at the expense of other advantages, i.e. if in the interest of great accuracy, use of more difficult average is desirable one should prefer that.

(iii) It should be based on all the observations: The average should depend upon each and every observation so that if any of the observation is dropped average itself is altered.

(iv) It should be rigidly defined: An average should be properly defined so that it has one and only one interpretation. It should preferably by defined by an algebraic formula so that if different people compute the average for the same figures they all get the same answer (barring arithmetical mistakes).

(v) It should be capable of further algebraic treatment: We should prefer to have an average that could be used for further statistical computation. For example, if we are given separately the figures of average income and number of employees of two or more factories, we should be able to compute the combined average.

(vi) It should not be unduly affected by the presence of extreme values: Although each and every observation should influence the value of the average, none of the observation should influence it unduly. If one or two very small or very large observations unduly affect the average i.e. either increase its value or reduce its value, the average cannot be really typical of the entire set of data. In the words, extremes may distort the average and reduce its usefulness.

The following are important measures of central tendency which are generally used in business:

3.1.1 Arithmetic Mean

The arithmetic mean (usually denoted by the symbol $\bar{x}$) of a set of observations is the value obtained by dividing the sum of all observations in a series by the number of items constituting the series.

Computation of Arithmetic Mean:

1. Un-grouped Data: If $x_1, x_2, \ldots\ldots\ldots\ldots x_n$ are the n given observations, then their arithmetic mean usually denoted by $\bar{x}$ is given by:

$$\bar{x} = \frac{X1 + X2 + X3 \ldots\ldots\ldots X4}{N}$$

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

The symbol Σ (Greek letter called Sigma) denotes the sum of n items. In normal use only Σ x is written in place of Σx (i=1...n). However, when the sum is combined to a given range of numbers out of the total, then it becomes necessary to specify. j

Problem 1:

The following gives the marks obtained by 10 students at an examination:

Roll Nos.:	1	2	3	4	5	6	7	8	9	10
Marks Obtained :	43	48	55	57	21	60	37	48	78	59

Calculate the arithmetic mean.

Solution: Computation of Arithmetic Mean

Roll No.	***Marks obtained (x)***
1	43
2	48
3	55
4	57
5	21
6	60
7	37
8	48
9	78
10	59
Total	Σx=506

Arithmetic Mean= $(\sum x)/n$

= 506/10

= 50.6

$\bar{x}$ = 50.6 Ans.

Frequency Distribution: In case of a frequency distribution. The arithmetic mean is given by the following formula:

$$= \frac{f_1x_1+f_2x_2+\ldots\ldots f_nx_n}{f_1+f_1+\ldots\ldots f_n} = \frac{\Sigma f_ix_i}{\Sigma f_i} = \frac{\Sigma f_ix_i}{N}$$

Where N=Σf is the total frequency. The mean value obtained in this manner is sometimes referred as *weighted arithmetic mean,* as distinct from *simple arithmetic mean.*

In case of continuous or grouped frequency distribution, the value of x is taken as the mid-value of the corresponding class.

Problem 2:

From the following data of marks obtained by 50 students of a class, calculated the arithmetic mean:

Marks	*No. of Students*	*Marks*	*No. of Students*
20	8	50	5
30	12	60	6
40	15	70	4

Let the marks be denoted by X and the number of students by f.

Solution:

Marks(x)	*No. of Students (f)*	*fx*
20	8	160
30	12	360
40	15	600
50	5	250
60	6	360
70	4	280
		2010

$$\bar{X} = \frac{\Sigma fx}{N} = \frac{2010}{50} = 40.2 \text{ marks}$$

Hence the average mark is 40.2.

Problem 3: Calculate the mean for the following frequency distribution:

Sales (in Rs. lakh):	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of firms :	6	5	8	15	7	6	3

Solution: Computation of Arithmetic Mean

Sales	***Mid-Value***	***No. of Firms***	***fX***
(in Rs. lakh)	***(X)***	***(f)***	
0-10	5	6	30
10-20	15	5	75
20-30	25	8	200

30-40	35	15	525
40-50	45	7	315
50-60	55	6	330
60-70	65	3	195
		Σf=50	ΣfX=167

$$A.M. = \frac{\Sigma fx}{N} = \frac{1670}{50} = Rs.\ 33.4\ lakhs.$$

Short-cut Method: When the short-cut method is used arithmetic mean is computed by applying the formula given below:

$$X = A + \frac{\Sigma fd}{N}$$

Where, A = assumed mean and d=deviations from assumed mean (m-A).

Problem 3 will be solved as follows when short-cut method is used :

Mid value (m) :	5	15	25	35	45	55	65
Deviations (m-A) :	-30	-20	-10	0	10	20	30
A = 35f :	6	5	8	15	7	6	3
fd :	-180	-100	-80	0	70	120	90

Here: A = 35, N = 50, Σ fd = -80,

$$\bar{x} = 35 + \frac{-80}{50}$$

= 33.4 lakhs.

Step-deviation Method: In the step deviation method the only additional point is that in order to simplify calculations we take a common factor from the data and multiply the result by the common factor. The formula is:

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

Where A = assumed mean; F = frequency; D' = $\frac{(m - A)}{(C)}$;

C= common factor, N = Total number of observations.

The step deviation method is most commonly used formula as it facilitate calculations.

Problem 4:

The following table gives the individual output of 180 female workers at a particular plant during a work. Find out the average output per worker.

Output (in units)	500-509	510-519	520-529	530-539
No. of workers	8	18	23	37
Output (in units)	540-549	550-559	560-569	570-579
No. of workers	47	26	16	6

Solution :

Mid-value (m)	**Frequency (*f*)**	**D_1=m-534.5 10**	***f*d'**
504.5	8	- 3	-24
514.5	18	- 2	-36
524.5	23	- 1	-23
534.5	37	0	0
544.5	47	1	47
554.5	26	2	52
564.5	16	3	48
574.5	5	4	20
	180		Σfd = 84

Average output : $\bar{X} = A + \frac{C}{N} \times \Sigma fd$

$$= 534.5 + \frac{10}{180} 84$$

$$= 534.5 + 4.67$$

$$= 539.17 \text{ units}$$

Mean of the Combined Series

If a group of n_1 observations has A.M. $\overline{X}_1$ and another group of n_2 observations has A.M. $\overline{X}_2$, then the A.M. ($\overline{X}_{12}$) of the composite group of $n_1 + n_2$ (=n, say) observations is given by

$$\overline{X}_{12} = \frac{n_1 \overline{X}_1 + n_2 X_2.}{n_1 + n_2}$$

where, $\overline{X}_{12}$ = combined mean of the two series or two groups of data. This can be generalised to any number of groups.

Problem 5 : The mean height of 25 male workers in a factory is 61 cms., and the mean height of 25 female workers in the same factory is 58 cms. Find the combined mean height of 50 workers in the

Solution :

$$X_{12} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$$

$$n_1 = 25, \overline{x}_1 = 61, n_2 = 25\ \overline{x}_2 = 58$$

$$\overline{x}_{12} = \frac{25x61 + 25x58}{50+50} = \frac{1525 + 1450}{50} = \frac{2975}{50} = 59.6$$

Thus combined mean height of 50 workers is 59.5 cms.

Merits and Limitations of Arithmetic Mean

The arithmetic mean is the most popular average in practice. It is due to the fact that it possesses first five out of six characteristics of a goods average (as discussed earlier) and no other average possesses such a large number of characteristics.

However, arithmetic mean is unduly affected by the presence of extreme values. Also in open-end frequency distribution it is difficult to compute mean without making assumption regarding the size of the class-interval of the open-end classes.

Mathematical Properties of Arithmetic Mean

The following area a few important mathematical properties of the arithmetic mean.

1. The sum of the deviations of the items form the arithmetic mean (taking signs into account) is always zero. i.e. Σ(x-x) = 0.
2. The sum of the squared deviations of the items from arithmetic mean is minimum, that is, less than the sum of the squared deviations of the items from any other value.
3. If we have the arithmetic mean and number of items of two or more than two related groups, we can compute combined average of these groups.

3.2.2 MEDIAN

In the words of L.R. Conner : "The median is that value of the variable which divides the data in two equal parts, one part comprising all the values greater and the other, all values less than median." Thus, as against arithmetic mean which is based on all the items of the distribution, the median is only positional average, i.e. the value depends on the position occupied by a value in the frequency distribution.

Computation of Median

1. **Ungrouped data :** If the number of observation is odd, then the median is the middle value after the observations have been arranged in ascending or descending order of magnitude. In case of even number of observations median is obtained as the arithmetic mean of two middle observations after they are arranged in ascending or descending order of magnitude.

Problem 6: The marks obtained by 12 students out of 50 are: 25, 20, 23, 32, 40, 27, 30, 25, 20, 10, 15, 41

Solution: The values obtained by 12 students arranged in ascending order as: 10, 15, 20, 20, 23, 25, 25, 27, 30, 32, 40, and 41

Here the number of items 'N' = 12, which is even

∴ The two middle items are 6th and 7th items

$$\text{i.e.} \quad \frac{25+25}{2} = 25 \text{ is the median value.}$$

2. **Frequency (Discrete) Distribution:**

In case of frequency distribution where the variables take the value $X_1, X_2, \ldots\ldots,$

ΣX_n with respective frequencies $f_1, f_2, ..., f_n$ with N= f, median is the size of the ½(N+1) th item or observation. In this case the use of cumulative frequency (c.f.) distribution facilitates the calculations. The steps involved are:

(i) Prepare the less than cumulative frequency (c.f.) distribution.

(ii) Find N/2.

(iii) Find the c.f. just greater than N/2.

(iv) The corresponding value gives the median.

Problem 7: From the following data find the value of median:

Income (Rs.)	1000	1500 800	2000	2100	1700
No. of Persons	24	26 14	10	5	28
Solution:					
Income arranged in ascending order		***No. of persons (f)***			***c.f.***
800		14			14
1000		24			38
1500		26			64
1700		28			92
2000		10			102
2100		5			107

Median = Size of (N/2)th item = $\frac{107}{2}$ = 53.5

53.5th item is consisted in the c.f. = 64. The corresponding value to this = 1500. Hence Median = Rs. 1500.

3. Continuous Frequency Distribution : Steps involved for its computation are :

(i) Prepare less than cumulative frequency (c.f.) distribution.

(ii) Find N/2.

(iii) Locate c.f. just greater than N/2.

(iv) The corresponding class contains the median value and is called the median class.

(v) The value of median is now obtained by using the interpolation formula :

$$\text{Median (Md)} = 1 + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

Where 1 is the lower limit or boundary of the median class; f is the frequency of the median class; h is the magnitude or width of class interval; n = Σf is the total frequency; and C is the cumulative frequency of the class preceding the median class.

Problem 8: The annual profits (in Rs. lacs) shown by 60 firms are given below:

Profits:	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65
No. of firms:	4	5	11	6	5	8	9	6	4	2

Calculate the median.

Solution :

Profits	***No.of firms (f)***	***Cumulative frequency (c.f.)***
15-20	4	4
20-25	5	9
25-30	11	20
30-35	6	26
35-40	5	31
40-45	8	39
45-50	9	48
50-55	6	54
55-60	4	58
60-65	2	60

Median item = $\frac{1}{2}$ N = 30

The cumulative frequency just greater than 30 is 31 and is corresponding class 35-40 is the median class.

$$\text{Median} = L + \frac{N/2 - c.f.}{f} \times h$$

$$= 35 + \frac{30-26}{5} \times 5 = 39 \text{ marks.}$$

Merits and Limitations of Median

The median is superior to arithmetic mean in certain aspects. For example, it is especially useful in case of open-ended distribution and also it is not influenced by the presence of extreme values. In fact when extreme values are present in a series, the median is more satisfactory measure of central tendency than the mean.

However, since median is positional average, its value is not determined by each and every observation. Also median is not capable of algebraic treatment. For example, median cannot be used for determining the combined median of two or more groups. Furthermore, the median tends to be rather unstable value if the number of observations is small.

3.2.3 MODE

Mode is the value which occurs most frequently in the set of observations.

Computation of Mode

(a) Ungrouped Data: In case of a frequency distribution, mode is the value of the variable corresponding to the maximum frequency.

Problem 9: Calculate the mode of the following data:

Sr. No.	*Marks obtained*	*Sr. No.*	*Marks obtained*
1	16	6	27
2	27	7	20
3	24	8	18
4	12	9	15
5	27	10	15

Solution. : Calculation of Mode

Size of item (Marks)	*No. of times it occurs*	*Size of item (Marks)*	*No. of times it occurs*

12	1	20	1
15	2	24	1
16	1	27	3
18	1		

Since the item 27 occurs the maximum number of times i.e. 3, hence the modal marks are 27.

(b) Grouped Data: From the grouped frequency distribution, it is relatively difficult to find the mode accurately. However, if all classes are of equal width, mode is usually calculated by the formula :

$$\text{Mode} = L + \frac{A_1}{A_1 + A_2} \times h$$

Where, L = the lower limit or boundary of the modal class;

h = magnitude or width of the modal class'

$A_1 = f_1 - f_0$, $A_2 = f_1 - f_2$;

f_1 = frequency of the modal class;

f_0 = frequency of the class preceding the modal class; and f_2 = frequency of the class succeeding the modal class.

Mode is generally abbreviated by the symbol M_0.

The above formula takes the following form:

$$\text{Mode (Mo)} = L + \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \times h = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h \qquad (1)$$

Problem 10 : Calculate mode from the following data :

Marks	*No. of students*	*Marks*	*No. of students*
Above 0	80	Above 60	28
Above 10	77	Above 70	16
Above 30	65	Above 80	10
Above 40	55	Above 90	8

Above 50	43	Above 100	0

Solution:

Since this is cumulative frequency distribution, we are to first convert it into a simple frequency distribution.

Marks	*No. of students*	*Marks*	*No. of students*
0-10	3	50-60	15
10-20	5	60-70	12
20-30	7	70-80	6
30-40	10	80-90	2
40-50	12	90-100	8

By inspection the modal class is 50-60.

$$\text{Mode} = L + \frac{A_1}{A_1 + A_2} \times i$$

$L = 50$, $A_1 = (15-12) = 3$, $A_2 = (15-12) = 3$, $i = 10$

$$M_0 = 50 + \frac{3}{3+3} \times 10 = 50+5 = 55 \text{ Marks.}$$

Empirical Relation between Mean (X), Median (Md) and Mode (M_0)

In case of a symmetrical distribution mean, median and mode coincide i.e. Mean = Median = Mode. However, for a moderately asymmetrical (non-symmetrical) distribution, mean and mode usually lie on the two ends and median lies in between they and they obey the following important empirical relationship, given by Prof. Karl Pearson.

Mode = 3 Median - 2 Mean --------------------------(2).

While applying the formula (1) for calculating mode, it is necessary to see that class intervals are uniform throughout. If they are unequal they should first be made equal on the assumption that the frequencies are equally distributed throughout the class, otherwise we will get misleading results.

A distribution having only one mode is called unimodel. If it contains more than one mode, it is called bimodal or multimodal. In the latter case the values of the mode cannot be determined by formula (1) and hence mode is ill-defined when there is more than one value of mode. Where mode is ill-defined, its value is ascertained by the formula (2) based upon the relationship between mean, median and mode. Mode = 3 Median - 2 Mean.

Merits and Limitations of Mode

Like Mean, the mode is not affected by extreme values and its value can be obtained in open-end distribution without ascertaining the class limits. Mode can be easily used to describe qualitative phenomenon. For example, when we want to compare the consumer preferences for different types of products, say, soap, toothpastes etc. or different media of advertising, we should compare the modal preferences. In such distributions where there is an outstanding large frequency, mode happens to be meaningful as an average.

However, mode is not rigidly defined measure as there are several formulae for calculating the mode, all of which usually give somewhat different answer. Also the value of mode cannot always be computed, such as, in case of binomial distributions.

3.3 GEOMETRIC MEAN

The Geometric mean (usually abbreviated as G.M.) of a set of n observations is the nth root of their product.

Computation of Geometric Mean

The Geometric Mean G.M. of n observations X_i, i=1, 2,……, n is G.M. = $(X_1.X_2.........X_n)^{1/n}$

The computation is facilitated by the use of logarithms. Taking logarithms of both sides, we get.

$$\log G.M. = \frac{1}{n}(\log x_1 + \log x_2 + ... + \log x_n) = \frac{1}{n}\sum_{i=1}^{n}\log x_i$$

$$G.M. = \text{Antilog}\left(\frac{1}{n}\sum_{i=1}^{n}\log x_i\right) \text{ or antilog}\left(\frac{1}{n}\Sigma \log\right)$$

Problem 11: From the data given below calculate the G.M.

15, 250, 15.7, 157, 1.57, 105.7, 10.5, 1.06, 25.7, 0.257

Solution:

Value (x)	***Log (x)***
15	1.1761

250	2.3979
15.7	1.1959
157	2.1959
1.57	0.1959
105.7	2.0240
10.5	1.0212
1.06	0.0253
25.7	1.4099
0.257	0.0409
Total	**11.0520**

$$G.M. = \text{Antilog}\left(\frac{1}{n}\Sigma \log x\right)$$

$$G.M. = \text{Antilog}\left(\frac{11.0520}{10}\right) = 12.75$$

In case of frequency distribution x_i/f_i (i=1, 2,.................... , n)geometric mean, G.M. is given by

$$G.M. = n\sqrt{(x_1.x_1 \ldots\ldots f_1 \text{ times})(x_2.x_2 \ldots\ldots f_2 \text{ times}) \ldots\ldots (x_n.x_n \ldots\ldots f_n \text{ times})}$$

Since the product of the values in a frequency distribution is usually very large, formula (3) is not suitable in computing the value of G.M. Taking logarithm of both sides in (3), we get :

$$\log G.M. = \frac{1}{N}\{\log(x_1^{f1}.x_2^{f2} \ldots\ldots x_n^{fn})\}$$

$$= \frac{1}{N}\{f_1 \log x_1 + f_2 \log x_2 + \ldots\ldots\ldots\ldots\ldots + f_n \log x_n)$$

Problem 12: Calculate Geometric Mean of the following distribution.

X	:	70	100	103	107	149	
f	:	10	12	8	5	5	

Solution :						
	X		***f***		***log x***	***f log x***
	70		10		1.8451	18.4512
	100		12		2.0000	24.0000
	103		8		2.0128	16.1024
	107		5		2.0294	10.1470
	149		5		2.1732	10.8690
						79.5664

$$\text{Log G.M.} = \frac{\Sigma f \log x}{\Sigma f}$$

$$= \frac{79.5664}{40}$$

$$= 1.989$$

In the case of grouped frequency distribution, the values of x are the mid-values of the corresponding classes.

Combined Geometric Mean

Just as we have talked of combined arithmetic mean, in a similar manner we can also talk of combined geometric mean. If the Geometric mean of N observations is G.M. and these observations are divided into two sets containing N_1 and second containing N_2 observations having GM_1 and GM_2 as the respective geometric means, then:

$N_1 \log GM_1 + N_2 \log GM_2$ log GM = ---------------------------------- $N_1 + N_2$

Merits and Limitations of Geometric Mean

Geometric mean is highly useful in averages, ratios, percentages and in determining rates of increase and decrease. It is also capable of algebraic manipulation. For example, if the geometric mean of two or more series and their number of observations are known, a combine geometric mean can easily be calculated. However, compared to arithmetic mean, this average is more difficult to compute and interpret. Also geometric mean cannot be computed when odd number of observations is negative.

3.4 HARMONIC MEAN

Harmonic mean of a number of observations is the reciprocal of arithmetic mean of reciprocals of the given values.

Computation of Harmonic Mean: If $X_1, X_2, \ldots\ldots, X_n$ are the n observations, their harmonic mean (abbreviated as H) is given by :

$$H.M.\ (H) = \frac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots + \frac{1}{x_n}\right)} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Problem 13: Find the Harmonic mean from the following:

2574, 475, 75, 5, 0.8, 0.08, 0.005, 0.0009

Solution:

X	*1/x*	*X*	*1/X*
2574	0.0004	0.8	1.2500
475	0.0021	0.08	12.5000
75	0.0133	0.005	200.0000
5	0.2000	0.0009	1111.1111

Σ(1/x)=1325.0769

$$H.M. = \frac{n}{(1/x)\ \Sigma} = \frac{8}{1325.0769} = 0.006$$

In case of frequency distribution, we have

$$\frac{1}{H} = \frac{1}{N}\left[\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots\ldots\ldots + \frac{f_n}{x_n}\right] \text{ where } N=\Sigma f$$

Problem 14: The following table gives weights of 31 persons in a sample enquiry. Calculate mean by using Harmonic mean.

Weight (in 1bs):	130	135	140	145	146	148	149	150	157
No. of persons:	3	4	6	6	3	5	2	1	1

Solution:

(Weight(x)	*Frequency(f)*	*1/x*	*f(1/x)*
130	3	.00769	.02307
135	4	.00741	.02964
140	6	.00714	.04284
145	6	.00690	.04140
146	3	.00685	.02055
148	5	.00676	.03380
149	2	.00671	.01342
150	1	.00667	.00667
157	1	.00637	.00637
	31		.21776

$$\frac{1}{\text{H.M.}} = \frac{\text{Sf. } 1/x}{N} = \frac{.21776}{31} = .007024$$

$$\text{Or H.M.} = \frac{1}{.007024} = 142.4 \text{ lbs.}$$

The harmonic mean is restricted in its field of application. The harmonic mean is a measure of central tendency for data expressed as rates, for instance - kms. per hour, tonnes per day, kms per litre etc.

Merits and Limitations of Harmonic Mean

The harmonic mean, like the arithmetic mean and geometric mean is computed from all observations. It is useful in special cases for averaging rates. However, harmonic mean gives largest weight to smallest observations and as such is not a good representation of a statistical series. In dealing with business problems harmonic mean is rarely used.

3.5 Check Your Progress

There are some activities to check your progress. Answer the followings:

1. The sum of the squared deviations of the items from arithmetic mean is…………….
2. Harmonic mean gives largest weight to ………………observations and as such is not a good

representation of a statistical series.

3. Geometric mean cannot be computed when odd number of observations is…………….
4. Mode can be easily used to describe ……………………..phenomenon.
5. In case of a ……………………..distribution mean, median and mode coincide.

3.6 Summary

It is the most important objective of statistical analysis is to get one single value that describes the characteristics of the entire mass of cumbersome data. Such a value is finding out, which is known as central value to serve our purpose. Central tendency or 'average' value is the powerful tool of analysis of data that represents the entire mass of data. An 'average' thus is a single value which is considered as the most representative or typical value for a given set of data. Such a value is neither the smallest nor the largest value, but is a number whose value is somewhere in the middle of the group. Measures of central tendency show the tendency of some central value around which the data tends to cluster. There are different measure for central tendency like mean, median, mode, harmonic mean and geometric mean. There are various merits and limitations of each and having different characteristics.

3.7 Keywords

Average: It is a single value which is considered as the most representative or typical value for a given set of data.

Mean: It a set of observations is the value obtained by dividing the sum of all observations in a series by the number of items constituting the series.

Median: It is that value of the variable which divides the data in two equal parts, one part comprising all the values greater and the other, all values less than median.

Mode: It is the value which occurs most frequently in the set of observations.

3.8 Self- Assessment Test

Q1. What is the measures of central tendency? Why are they called measures of central tendency?

Q2. What is the properties of a good average?

Q3. Give a brief note of the measures of central tendency together with their merits and demerits. Which is the best measured of central tendency and why?

Q4. Following distribution gives the pattern of overtime work done by 100 employees of a company. Calculate median.

Overtime Hours:	10-15	15-20	20-25	25-30	30-35	35-40

No. of Employees: 11 20 35 20 8 6

Q5. The mean monthly salary paid to all employees in a company is Rs. 1600. The mean monthly salaries paid to technical and non-technical employees are Rs. 1800 and Rs. 1200 respectively. Determine the percentage of technical and non-technical employees of the company.

Q6. Calculate the arithmetic mean and the median of the frequency distribution given below. Also calculate the mode using the empirical relation among the three:

Class Limits	***Frequency***	***Class Limits***	***Frequency***
130-134	5	150-154	17
135-139	15	155-159	10
140-144	28	160-164	1
145-149	24		

Q7. In a certain factory a unit of work is completed by A in 4 minutes, by B in 5 minutes, By C in 6 minutes, by D in 10 minutes and by E in 12 minutes.

(a) What is the average number of units of work completed per minute?

(b) At this rate how many units will they complete in a six-hour day?

Q8. Find the average rate of increase in population which in the first decade increased by 20%, in the second decade by 30% and in the third decade by 40%.

Q9. In a class of 50 students, 10 has failed and their average of marks is 2.5. The total marks secured by the entire class were 281. Find the average marks of those whohave passed.

3.9 Answers to Check your Progress

1. Minimum
2. Smallest
3. Negative
4. Qualitative
5. Symmetrical

3.10 References/Suggested Readings:

- Gupta, S. P.: Statistical Methods, Sultan Chand and Sons, New Delhi.
- Levin, R. I. and David, S. R.: Statistics for Management, Prentice Hall, New Delhi.
- Gupta, C. B.: Introduction to Statistical Methods.

- Hooda, R. P.: Statistics for Business and Economics, Macmillan, New Delhi.

Subject: Business Statistics-1	
Course Code: 302	**Author :** Dr. Pradeep Gupta
Lesson No. : 4	**Vetter:** Dr. B.S. Bodla
MEASURES OF DISPERSION	

Structure

4.0 LEARNING OBJECTIVES

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

- Understand the concept of dispersion
- Understand the concept of Range
- Understand the concept of Quartile Deviation

- Understand the concept of Mean Deviation
- Understand the concept of Standard Deviation
- Understand the concept of coefficient of Variance
- Explain the concept of Lorenz Curve

4.1 INTRODUCTION

The value given by a measure of central tendency is considered to be the representative of the whole data. However, it can describe only one of the important characteristics of a series. It does not give the spread or range over which the data are scattered. Measures of dispersion are used to indicate this spread and the manner in which data are scattered.

4.1.1 DEFINITION

Dispersion indicates the measure of the extent to which individual items differ from some central value. It indicates lack of uniformity in the size of items. Some important definitions of dispersion are given below:

(1) According to Spiegel, "The degree to which numerical data tend to spread about an average value is called the variation of dispersion of the data."

(2) Simpson and Kafka define dispersion as "The measurement of the scatterness of the mass of figures in a series about an average is called measure of variation or dispersion."

(3) As defined by Brooks and Dick, "Dispersion or spread is the degree of the scatter or variation of the variable about a central value."

Since measures of dispersion give an average of the differences of various items from an average, they are also called averages of the *second order.*

4.1.2 USES OF MEASURES OF DISPERSION

Average is a typical value but it alone does not describe the data fully. It does not tell us how the items in a series are scattered around it. To clear this point considers the following three sets of data:

Set A	30	30	30	30	30
Set B	28	29	30	31	32
Set C	3	5	30	37	75

All the three sets A, B and C have mean 30 and median is also 30. But by inspection it is apparent that the three sets differ remarkably from one another. Thus to have a clear picture of data, one needs to have a measure of dispersion or variability (scatteredness) amongst observations in the set. It is also used for comparing the variability or consistency (uniformity) of two or more series. A higher degree of variation means smaller consistency.

4.1.3 PROPERTIES OF A GOOD MEASURE OF DISPERSION

There are various measures of dispersion. The difficulty lies in choosing the best measure as no hard and fast rules have been made to select any one. However, some norms have been set which work as guidelines for choosing a particular measure of dispersion. A measure of dispersion is good or satisfactory if it possesses the following characteristics.

- It is easily understandable.
- It utilizes all the data.
- It can be calculated with reasonable ease and rapidity.
- It affords a good standard of comparison.
- It is suitable for algebraic and arithmetical manipulation.
- It is not affected by sampling variations.
- It is not affected by the extreme values.

4.1.4 VARIOUS MEASURES OF DISPERSION

Commonly used measures of dispersion are:

(a) Range

(b) Quartile deviation

(c) Mean deviation

(d) Standard deviation

(a) Range

Definition. Range is the difference between the two extreme items, i.e. it is the difference between the maximum value and minimum value in a series.

Range (R) = Largest value (L) minus Smallest value (S) A relative measure known as *coefficient of range is* given as:

$$\text{Coefficient of range} = \frac{L - S}{L + S}$$

Lesser the range or coefficient of range, lower the variability.

Properties.

(a) It is the simplest measure and can easily be understood.

(b) Besides the above merit, it hardly satisfies any property of a good measure of dispersion e.g. it is based on two extreme values only, ignoring the others. It is not liable to further algebraic treatment.

Example 1. The population (in '000) in eighteen Panchayat Samities of a district is as given below:

77, 76, 83,	68,	57,	107, 80,	75,	95,	100, 113, 119,
121, 121, 83,	87,	46,	74			

Calculate the range and coefficient of range.

Solution. Largest value (L) Smallest value (S)		= 121 = 46
Range (R)		= L - S
		= 121-46 = 75
Coefficient of range = $\frac{L - S}{L + S}$	=	$\frac{121-46}{121+46} = \frac{75}{167} = 0.449$

Range for grouped data. In case of grouped data, the range is the difference between the upper limit of the highest class and the lower limit of the lowest class. No consideration is given to frequencies.

Example 2. Find range of the following distribution.

Class-interval	45-49	50-54	55-59	60-64	65-69
Frequency	37	26	8	5	1

Solution. The series can be written as follows :

Group	Frequency
44.5-49.5	37
49.5-54.5	26
54.5-59.5	8
59.5-64.5	5
64.5-69.5	1
Range = 69.5-44.5 = 25	

(b) Quartile Deviation

Quartile deviation is obtained by dividing the difference between the upper quartile and the lower quartile by 2.

$$\text{Quartile deviation or Q.D.} = \frac{\text{Upper Quartile - Lower Quartile}}{2}$$

$$= \frac{Q_3 - Q_1}{2}$$

The coefficient of quartile deviation is given by the following formula:

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 - Q_1}$$

Coefficient of quartile deviation is a relative measure of dispersion and is used to compare the variability among the middle 50 per cent observations.

Properties.

(i) It is better measure of dispersion than range in the sense that it is based on the middle 50 per cent observations of a series of data rather than only two extreme values of a series.

(ii) It excludes the lowest and the highest 25% values.

(iii) It is not affected by the extreme values.

(iv) It can be calculated for the grouped data with open end intervals.

(v) It is not capable of further algebraic treatment.

(vi) It is not considered a good measure of dispersion as it does not show the scattering of the central value. In fact it is a measure of partitioning of distribution. Hence it is not commonly used.

Example 3. Given the number of families in a locality according to monthly per capita expenditure classes in rupees as:

Class-interval	140-150	150-160	160-170	170-180	180-190	190-200	200-210	210-220	220-230	230-240	240-250
No. of families	17	29	42	72	84	107	49	34	31	16	12

Calculate Quartile deviation and coefficient of quartile deviation.

Solution.

Monthly per capita expenditure (Rs.)	Number of Families (f)	Cumulative frequency (c.f.)
140-150	17	17
150-160	29	46
160-170	42	88
170-180	72	160
180-190	84	244
190-200	107	351
200-210	49	400
210-220	34	434
220-230	31	465
230-240	16	481
240-250	12	493

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

(i) To calculate Q_1, we have to first find :

$$\frac{N}{4} = \frac{493}{4} = 123.25$$

The number 123.25 is contained in the cumulative frequency 160. Hence the first quartile lies in the class 170-180. By using the formula for Q_1 we get,

$$Q_1 = L + \frac{N/4 - c.f.}{f} * i$$

L = 170, N/4 = 123.25, c.f. = 88, f=72, i=10

$$Q_1 = 170 + \frac{123.25 - 88}{72*10}$$

= Rs. 174.90

(ii) To calculate Q_3 we find:

$$\frac{3N}{4} \qquad \frac{3 \times 493}{4}$$

$= 369.75$

The number 369.75 is contained in the cumulative frequency 400. Hence the class 200-210 is the third quartile class. By using the formula for Q_3 we get:

$Q_3 = L + (3N/4 - c.f.)/f * i$
L= 200, 3 N/4 = 369.75, c.f. = 351, f=49, i=10
$Q_3 = 200 + (369.75-351)/49 * 10$

= Rs. 203.83

Quartile Deviation (Q.D.) = (203.83-174.90)/2=28.93/2=14.465

Coefficient of Q.D = (203.83-174.90)/ (203.83+174.90)=28.93/378.73=0.076

(C) Mean Deviation

Mean deviation is the mean of deviations of the items from an average (mean, median or mode). As we are concerned with the deviations of the different values from an average and in finding the mean of deviations, we have to find the sum of deviations (whether positive or negative); we take all the deviations as positive. We are concerned with the deviations and not with their algebraic signs. We ignore negative signs because the algebraic sum of the deviations of individual values from the average is zero.

Calculation of mean deviation (M.D.).

Mean deviation of a set of n observations $x_1, x_2, ..., x_n$ is calculated as follows:

$$M.D. = \frac{1}{n} \sum_{i=1}^{n} |x_i - A|$$

for i= 1,2, ..., n where A is a central value.

Let $|x_i - A| = d_i$

$$\text{Then M.D.} = \frac{1}{n} \sum_{i=1}^{n} |d_i| \qquad ....(i)$$

In case data is given in the form of a frequency distribution, the variate values x_1 , x_2,......., x_n occur f_1, f_2, , f_n times respectively.

In such series the formula for mean deviation is,

$$M.D. = \frac{1}{N} \sum_{i=1}^{n} f_i \left| x_i - A \right| \quad \text{............ (ii)}$$

Where, N = Σ fi for i = 1, 2, ..., n

In case of grouped data, the mid-point of each class interval is treated as x_i and we can use the formula (ii) given above.

Properties.

(i) Mean deviation removes one main objection of the earlier measures, that it involves each value of the set.

(ii) Its main drawback is that algebraic negative signs of the deviations are ignored which is mathematically unsound.

(iii) Mean deviation is minimum when the deviations are taken from median.

(iv) It is not suitable for algebraic treatment.

Example. 4 :

Calculate mean deviation from the mean for the following data:

Size (x) :	2	4	6	8	10	12	14	16
Frequency :	2	2	4	5	3	2	1	1

Solution :

X	*f*	*F x*	$\|x-8\|$ $\|D\|$	$f\|d\|$
2	2	4	6	12
4	2	8	4	8
6	4	24	2	8
8	5	40	0	0
10	3	30	2	6

12	2	24	4	8
14	1	14	6	6
16	1	16	8	8
	N=20	Σ fx=1600		Σf|D|= 56

$$X = \frac{\Sigma fx}{N} = \frac{160}{20} = 8$$

$$M.D. = \frac{\Sigma f\ |D|}{N} = \frac{56}{20} = 2.8$$

Examples 5. Calculate the mean deviation (using median) from the following data.

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

Solution:

Size	Frequency (f)	Cummulative frequency	Deviation from median 9 |d|	f|d|
6	3	3	3	9
7	6	9	2	12
8	9	18	1	9
9	13	31	0	0
10	8	39	1	8
11	5	44	2	10
12	4	48	3	12
				Σ f |d| = 60

$$\text{Median} = \text{Size of } \frac{48+1}{2} \text{ th item}$$

$$= \text{Size of } 24.5^{th} \text{ item} = 9$$

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

Example. 6 : Find the median and mean deviation of the following data :

Size	*Frequency*	*Size*	*Frequency*
0-10	7	40-50	16
10-20	12	50-60	14
20-30	18	60-70	8
30-40	25		

Solution : Calculation of Median and Mean Deviation.

Size	*f*	*c.f.*	*m.p.*	*|m-35.2|* *|D|*	*f |D|*
0-10	7	7	5	30.2	211.4
10-20	12	19	15	20.2	242.4
20-30	18	37	25	10.2	183.6
30-40	25	62	35	0.2	5.0
40-50	16	78	45	9.8	156.8
50-60	14	92	55	19.8	277.2
60-70	8	100	65	29.8	238.4
	N = 100			Σ f D = 1314.8 $\simeq$ 1315	

Median = Size of N/2 th item = 100/2 = 50th item.

Median lies in the class 30-40.

$$\text{Med.} = L + \frac{N/2 - c.f.}{f} \times i$$

L = 30, N/2 = 50, c.f. = 37, f= 25, i = 10.

$$\text{Med.} = 30 + \frac{50-37}{25} X\ 10 = 30 + 5.2 = 35.2$$

$$\text{M.D.} = \frac{\Sigma f D}{N} = \frac{1315}{100} = 13.15$$

Uses of Mean Deviation:

The outstanding advantage of the average deviation is its relative simplicity. It is simple to understand and easy to compute. Anyone familiar with the concept of the average can readily appreciate the meaning of the average deviation. If a situation requires a measure of dispersion that will be presented to the general public or any group not familiar with statistics, the average deviation is useful.

(D) Standard deviation

It is the square root of the quotient obtained by dividing the sum of squares of deviations of items from the Arithmetic mean by the number of observations.

$$\therefore \text{Standard deviation or } \sigma = \sqrt{\frac{\text{Sum of squares of deviation from A.M.}}{\text{Number of observations}}}$$

Standard deviation is an absolute measure of dispersion.

Calculation of standard deviation (σ).

(a) Ungrouped data

First method : $S.D.\ (\sigma) = \sqrt{\frac{\Sigma d^2}{n}}$

Where d is the deviation of value from the mean.

Second method : In this method, we assume a provisional mean and find the deviations of the values from the provisional mean. The following formula is applied under this method :

$$S.D.\ \text{or}\ \sigma = \sqrt{\frac{\Sigma d_x^2}{n} - \left[\frac{\Sigma d_x}{n}\right]^2}$$

Where d is the deviation of values of x observations from the assumed mean. This formula is more useful when values are in decimals and the mean of the series does not happen to be an integer.

In case the frequencies are also given, then standared deviation is calculated by using the formula :

$$S.D.\ \text{or}\ \sigma = \sqrt{\frac{\Sigma fd_x^2}{n} - \left[\frac{\Sigma fd_x}{n}\right]^2} \qquad \text{Where } n = \Sigma f$$

Example 7. Compute the standard deviation by the short method for the following data :

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

Solution. Let us assume that mean is 15.

X	d (x-15)	d^2
11	- 4	16
12	-3	9
13	-2	4
14	-1	1
15	0	0
16	1	1
17	2	4

18	3	9
19	4	16
20	5	25
21	6	36
	Σ d=11	Σ d2 = 121

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

$$= \sqrt{\frac{121}{11} - \left[\frac{11}{11}\right]^2}$$

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

$$= \sqrt{10} = 3.16$$

In continuous series, we take the central values of the groups.]

Example 8. : Find the standard deviation of the following distribution :

Age	:	20-25	25-30	30-35	35-40	40-45	45-50
No. of Persons	:	170	110	80	45	40	35

Take assumed average = 32.5

Solution :

Calculation of Standard Deviation.

Age	***m.p.***	***No. of persons***	***(m-32.5)/5***		
	M	***f***	***d***	***fd***	***fd^2***
20-25	22.5	170	- 2	-340	680
25-30	27.5	110	- 1	-110	110
30-35	32.5	80	0	0	0
35-40	37.5	45	1	45	45
40-45	42.5	40	2	80	160
45-50	47.5	35	3	105	315

	N=480	$\Sigma fd = -220$ $\Sigma fd^2 = 1310$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} \quad \left[\frac{\Sigma fd}{N}\right]^2} \times i$$

$$= \sqrt{\frac{1310}{480} \quad \left[\frac{-220}{480}\right]^2} \times 5$$

$$= \sqrt{2.729-.21} \times 5 \quad = 1.587 \times 5 = 7.936$$

Uses of the Standard deviation: As a measure of dispersion, standard deviation is most important. By comparing the standard deviations of two or more series, we can compare the degree of variability or consistency. It is a keystone in sampling and correlation and is also used in the interpretation of normal and skewed curves. It is used to gauge the representativeness of the mean also.

Merits of Standard Deviation:

- It is suitable for algebraic manipulation.
- It is less erratic.
- Standard deviation is considered to be the best measure of dispersion and is used widely.

Demerits of Standard Deviation:

- Its calculation demand greater time and labour.
- If the unit of measurement of variables of two series is not the same, then their variability cannot be compared by comparing the values of standard deviation.
- It gives more weight to extreme items and less to those which are nearer the mean. It is because of the fact that the squares of the deviations which are big in size would be proportionately greater than the squares of those deviations which are comparatively small. The deviations 2 and 8 are in the ratio of 1: 4 but their squares i.e. 4 and 64, would be in the ratio of 1: 16.

Mathematical Properties of Standard Deviation

Standard deviation has some very important mathematical properties which considerably enhance its utility in statistical work.

1. *Combined Standard Deviation :* Just as it is possible to compute combined mean of

two or more than two groups, similarly we can also compute combined standard deviation of two or more groups.

2. *Standard deviation of n natural numbers :* The standard deviation of the first in natural numbers can be obtained by the following formula :

$$\sigma = \frac{1}{12}\ (n^2-1)$$

3. The sum of the squares of deviations of items in the series from their arithmetic mean is minimum. This is the reason why standard deviation is always computed from the arithmetic mean.

4. The standard deviation enables us to determine, with a great deal of accuracy, where the values of a frequency distribution are located with the help of Teheycheff 's theorem, given by mathematician P.L. Tehebycheff (1821-1894). No matter what the shape of the distribution is, at least 75 percent of the values will fall within ± 2 standard deviation from the mean of the distribution, and at least 89 percent of values will be within ± 3 standard deviations from the mean.

   For a symmetrical distribution, the following relationships hold good:

   Mean ± 1 o covers 68.27% of the items. Mean ± 2 o covers 95.45% of the items. Mean ± 3 o covers 99.73% of the items.

4.2 Variance

The variance is just the square of the standard deviation value:

Variance = o^2 = $(S.D.)^2$

In a frequency distribution where deviations are taken from assumed mean, variance may directly be computed as follows

$$\text{Variance} = \left\{ \frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2 \right. \times i$$

Where $d = \frac{x-A}{i}$ and i = common factor.

Properties :

(i) The main demerit of variance is that its value is the square of the unit of measurement of variate values. For example, the variable x is measured in cms, the unit of variance is cm. Generally, this value is large and makes it difficult to decide about the magnitude of variation.

(ii) The variance gives more weightage to the extreme values as compared to those which are near to mean value, because the difference is squared in variance.

(iii) The combined groups without redoing the entire calculations.

(iv) Obviously, the combined standard deviation can be found by taking the square root of the combined variance.

Pooled or combined variance : By the combined variance of two groups, we mean the variance of the observations of the two groups taken together. Let us consider two groups consisting of n_1 and n_2 observations respectively. Suppose the means of the groups are $\overline{x}_1$ and $\overline{x}_2$ and the variances are σ_1^2 and σ_2^2 respectively. We know that the pooled mean of both the groups is,

$$\overline{X}_{12} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$$

The combined variance of the two groups is given by the formula :

$$\sigma_{12}^2 = \frac{n_1\ (\sigma_1^2 + d_1^2)\ +\ n_2\ (\sigma_2^2\ +\ d_2^2)}{n_1 + n_2}$$

Where, $d_1 = (\overline{x}_1 - \overline{x}_{12})$ and $d_2 = (\overline{x}_2 - \overline{x}_{12})$

The advantage of the formula of combined variance is that once we know the individual mean and variance of each group, we can calculate the variance of

Example 9: For a group of 50 male workers, the mean and standard deviation of their weekly wages are Rs. 63 and Rs. 9 respectively. For a group of 40 female workers these are Rs. 54 and Rs. 6, respectively. Find the standard deviation for the combined group of 90 workers.

Solution:

The data is $n_1 = 50 \quad \bar{x}_1 = 63 \quad \sigma_1 = 9$

$n_1 = 40 \quad \bar{x}_2 = 54 \quad \sigma_2 = 6$

Combined mean $\bar{x}_{12}$ for group of 90 $= \dfrac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$

$= [50 \times 63 + 40 \times 54] / 90$

$= [3150 + 2160]/90 = 5310/90 = 59$

Combined standard deviation $= \dfrac{n_1 (\sigma_1^2 + d_1^2) + n_2 (\sigma_2^2 + d_2^2)}{n_1 + n_2}$

Where, $d_1 = (\bar{x}_1 - \bar{x}_{12})$ and $d_2 = (\bar{x}_2 - \bar{x}_{12})$

$\sigma^2 12 = [50 (81+16) + 40 (36+25)] / 90$

$= [97 \times 50 + 40 \times 61] / 90 = [4850 + 2440] / 90$

$= 7290 / 90 = 81$

$\therefore \sigma_{12} = 9$

Example 10 : The analysis of the results of a budget survey of 150 families gave an average monthly expenditure of Rs. 120 on food items with a standard deviation of Rs. 15. After the analysis was completed it was noted that the figure recorded for one household was wrongly taken as Rs. 15 instead of Rs.105. Determine the correct value of the average expenditure and its standard deviation.

Solution.

A.M. (x) = Rs. 120, No. of items = 150 Total as obtained = 120 x 150 = 18000

Correct total = (Total obtained - item misread) + correct item

= (18000-15) + 105 =18,090

Correct mean = Correct total/No. of items = 18090/150

= Rs. 120.6

$$(S.D.)^2 = \frac{\Sigma x2}{n} - \left[\frac{\Sigma x}{n}\right]^2$$

Before Correction $(15)^2 = \frac{\sum x^2}{150} - (120)^2$

Or $225 = \frac{\sum x^2}{150} - 14400$ or $\frac{\sum x2}{150} = 14,625$

$\sum x^2 = 14,625 \times 150$

Correct sum of squares = Sum of suqares before correction, minus square of misread item plus square of correct item

= 14625x150 - 15 x 15 + 105 x 105

= 15 x 15 [9750-1 + 7x7] = 225

= [9798]

Correct $(S.D.)^2$ = $225 \times 9798/150 - (120.6)^2$

	=	1.5 x 9798 - 14544.36
	=	152.64
Correct S.D.	=	12.4

4.2.1 Coefficient of variation (C.V.)

It two series differ in their units of measurement; their variability cannot be compared by any measure given so far. Hence in situations where either the two series have different units of measurements, or their means differ sufficiently in size, the coefficient of variation should be used as a measure of dispersion. It is a unit less measure of dispersion and also takes into account the size of the means of the two series. It is the best measure to compare the variability of two series or set of observations. A series with less coefficient of variation is considered more consistent.

Definition.

Coefficient of variation of a series of variate values is the ratio of the standard deviation to the mean multiplied by 100. If σ is the standard deviation and x is the mean of the set of values, the coefficient of variation is,

$$C.V. = \frac{\sigma}{\overline{x}} X 100$$

This measure was given by Professor Karl Pearson.

Properties:

(i) It is one of the most widely used measures of dispersion because of its virtues.

(ii) Smaller the value of C.V., more consistent is the data and vice-versa.

Hence a series with smaller C.V. than the C.V. of other series is more consistent, i.e. it possesses variability.

Example 11: A time study was conducted in a factory with the help of two samples A and B consisting of 10 workers. The time taken by the workers in each case recorded. From the particulars given below state which of the samples is more variable and which takes less time on an average. Time taken in minutes.

Sample A	130	125	120	135	140	145	130	145	140	150
Sample B	132	146	137	145	130	125	138	140	143	144

Solution : Let us calculate mean and standard deviation first by rearranging the data in ascending order.

For Sample A For Sample B

X	$d = \frac{x - 140}{5}$	d^2	y	d'= (y-140)	d'^2
120	- 4	16	125	-15	225
125	- 3	9	130	-10	100
130	- 2	4	132	- 8	64
130	- 2	4	137	- 3	9
135	- 1	1	138	- 2	4
140	0	0	140	0	0
140	0	0	143	3	9

145	1	1	144	4	16
145	1	1	145	5	25
150	2	4	146	6	36
	Σd=-8	Σd2=40		Σ d' = -20	Σ d'2 = 488

$$0 = 140 + \left(\frac{-8}{10} \times 5\right) = 136 \qquad y = 140 + \left(\frac{-20}{10}\right) = 138$$

This shows that on an average workers from sample A takes less time.

$$\sigma_x^2 = \left\{\frac{\Sigma d2}{n} - \left(\frac{\Sigma d}{n}\right)^2\right\} \times i^2 \text{ where } i = 5$$

$$= 25 \left\{\frac{40}{10} - \left(\frac{8}{10}\right)^2\right\} = 25 \left\{4 - \frac{16}{25}\right\} = 84$$

$$\therefore CV_x = \frac{\sigma x}{\overline{X}} \times 100 = \frac{\sqrt{84}}{136} \times 100 = 6.75\ \%$$

Similarly,

$$\sigma_y^2 = \left\{\frac{\Sigma d'2}{n} \quad \left(\frac{\Sigma d'}{n}\right)^2\right\}$$

$$= \left\{\frac{488}{10} - \left(\frac{-20}{10}\right)^2\right\}$$

$$= 48.8 - 4 = 44.8$$

$$\therefore CV_y = \frac{\sigma y}{\overline{y}} \times 100 = \frac{44.8}{138} \times 100 = 4.85\ \%$$

Since $CV_x > CV_y$, the sample 'A' is more variable, though as we have seen the workers of sample A takes less time.

4.2.2 RELATION BETWEEN STANDARD DEVIATION, MEAN DEVIATION AND QUARTILE DEVIATION

In any bell-shaped distribution, the S.D. will always be larger than M.D. and M.D. larger than Q.D. if the distribution approximates the form of normal curve, the M.D. will be 4/5 of S.D. and Q.D. will be about 2/3rd as large as S.D. Usually,

M.D. = 4/5 of S.D.

Q.D. = 2/3 of S.D.

4.2.3 COMPARISION OF THE VARIOUS MEASURES OF DISPERSION

Range is not a very satisfactory measure of dispersion because it depends solely on the two extreme values and may be very misleading if there are one or two abnormal items. It is impossible to estimate the range in case where there are open ends series. Therefore, it is an unreliable measure of dispersion. Quartile deviation is most easy to calculate and interpret but it is not amenable to mathematical treatment. Mean deviation is easy to compute but grouped data may be difficult. In almost all other aspects, the advantage rests with the standard deviation. Only the S.D. is suitable for algebraic manipulation. For this reason, it is used in correlation, in sampling and in other aspects of advanced statistics. Wecan compute the S.D. of the whole group if means and standard deviations of two or more subgroups are known. When it is required to compare two or more than two series or distributions, we compute relative measure of dispersion.

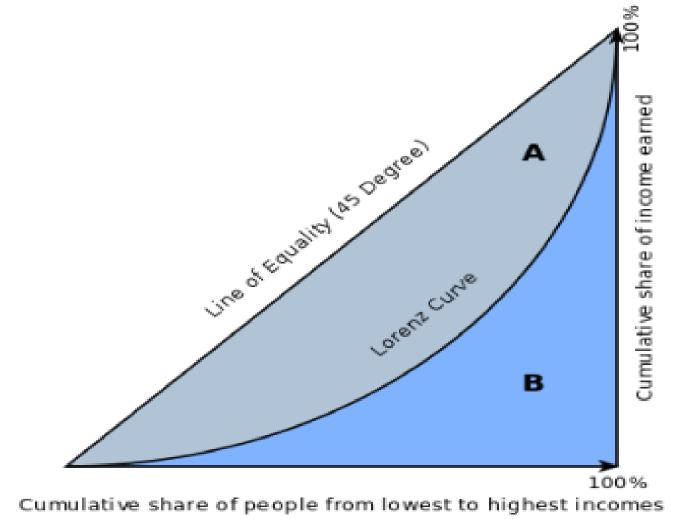
4.3 Lorenz Curve

A Lorenz curve is a graph used in economics to show **inequality in income spread or wealth**. It was developed by Max Lorenz in 1905, and is primarily used in economics. However, it may also be used to show inequality in other systems. The Gini index (The **Gini coefficient** is a statistic which quantifies the amount of inequality that exists in a population. The Gini coefficient is a number between 0 and 1, with 0 representing perfect equality and 1 perfect inequality) can be calculated from a Lorenz curve by taking the integral of the curve and subtracting from 0.5.

Reading a Lorenz curve

The x-axis on a Lorenz curve typically shows the portion or percentage of the total population and the y-axis shows the portion of total income/ wealth, or whatever is being analyzed. Since perfect equality would mean that a 1/k portion of the population controlled 1/k of the wealth, perfect equality on this graph would be shown by a straight line with a slope of 1. This line is often drawn on the graph as a point of reference, alongside the curved line which represents the actual wealth/income/size distribution. The further away from the 1/1 baseline a particular curve is, the more pronounced the inequality. Any point on the curve can be read to tell us what percentage or portion of the population command what percent of the wealth, income, or whatever variable is being studied. For instance, if the Lorenz curve representing income in a particular town crossed the point 0.57, 0.23 we would know that 0.57 of the population commanded just 0.23 of the town's income. In a completely equal situation, of course, 0.57 of the population would earn 0.57 of the total income, and the Lorenz curve would be identical to the 45 degree 1/1 line.

Both x and y axes are from 0 to 1, which can be expressed as a percentile (1 to 100 %) as shown in the above graph. The axes can also show quartiles.



Graphing a Lorenz Curve

To graph a Lorenz curve, the response variable (usually income or wealth) is first indexed in either equal or increasing order. Then points are graphed for a continuous distribution. If *n* is the number of instances of the response variable, then the ith x-coordinate will be i/n. The y-coordinate will be where Y_K is the response variables.

$$\frac{\sum_{k=1}^{i} Y_k}{\sum_{k=1}^{n} Y_k}$$

4.4 Check Your Progress

There are some activities to check your progress. Fill in the blanks:

(a) The variance is just the ……………………of the standard deviation value.

(b) Range is difference between the two............................

(c) Coefficient of variation is a ………………………measure of dispersion.

(d) Lorenz curve was developed by ………………………. in 1905.

(e) The outstanding advantage of the average deviation is its relative……………..

4.5 Summary

The value given by a measure of central tendency is considered to be the representative of the whole data. It does not give the spread or range over which the data are scattered. Measures of dispersion are used to indicate this spread and the manner in which data are scattered. Commonly used measures of dispersion are: Range, Quartile deviation, Mean deviation and Standard deviation. Range is the difference between the two extreme items, i.e. it is the difference between the maximum value and minimum value in a series. Quartile deviation is obtained by dividing the difference between the upper quartile and the lower quartile by 2. Mean deviation is the mean of deviations of the items from an average (mean, median or mode). Standard Deviation is the square root of the quotient obtained by dividing the sum of squares of deviations of items from the Arithmetic mean by the number of observations. The variance is just the square of the standard deviation value. If two series differ in their units of measurement; their variability cannot be compared by any measure given so far. Hence in situations where either the two series have different units of measurements, or their means differ sufficiently in size, the coefficient of variation should be used as a measure of dispersion. It is a unit less measure of dispersion and also takes into account the size of the means of the two series. It is the best measure to compare the variability of two series or set of observations. A series with less coefficient of variation is considered more consistent. Coefficient of variation of a series of variate values is the ratio of the standard deviation to the mean multiplied by 100. A Lorenz curve is a graph used in economics to show **inequality in income spread or wealth**.

4.6 Keywords

Range: It is the difference between the maximum value and minimum value in a series.

Quartile deviation: It is obtained by dividing the difference between the upper quartile and the lower quartile by 2.

Mean deviation: It is the mean of deviations of the items from an average (mean, median or mode).

Standard Deviation: It is the square root of the quotient obtained by dividing the sum of squares of deviations of items from the Arithmetic mean by the number of observations.

Variance: It is just the square of the standard deviation value.

Coefficient of variation: It is the ratio of the standard deviation to the mean multiplied by 100.

Lorenz Curve: A Lorenz curve is a graph used in economics to show **inequality in income spread or**

wealth.

4.7 Self-Assessment Questions

Q1. What does dispersion indicates about the data? Why is this of great importance?

Q2. What is the requirements of a good measure of dispersion?

Q3. Define and discuss the following terms.

- Quartile Deviation
- Mean Deviation
- Variance
- Coefficient of Variation

Q4. Calculate mean deviation from median as well as arithmetic mean.

Class-intervals	2-4	4-6	6-8	8-10
Frequencies	3	4	2	1

Q5. Calculate the standard deviation

Age	50-55	45-50	40-45	35-40	30-35	25-30	20-25
No.	25	30	40	45	80	110	170

Q6. Which of the two students was more consistent?

x	58	59	60	54	65	66	52	75	69	52
y	84	56	92	65	86	78	44	54	78	68

Q7. Calculate the appropriate measure of dispersion.

Wages in rupees	No. of wage earners
Less than 35	14
35-37	62
38-40	99
41-43	18
Over 43	7

Q8. In a certain distribution with n = 25 on measurements it was found that X=56 and o=2. After these results were computed it was discovered that a mistake had been made in one of the measurements which was recorded as 64. Find the mean and standard deviation if the incorrect value 64 is omitted.

Q9. The following table gives the frequency distribution of marks obtained by students of two classes. Find the arithmetic mean, the standard deviation and coefficient of variation of the marks of two classes.

Interpret the results.

Range of Marks	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
Class A	1	10	20	8	6	3	1	0
Class B	5	6	15	10	5	4	2	2

4.8 Answers to Check Your Progress

(a) Square

(b) Extreme items

(c) Unit less

(d) Max Lorenz

(e) Simplicity

4.9 References/Suggested Readings

1. Gupta, S. P.: Statistical Methods, Sultan Chand and Sons, New Delhi.
2. Levin, R. I. and David, S. R.: Statistics for Management, Prentice Hall, New Delhi.
3. Gupta, C. B.: Introduction to Statistical Methods.

Subject: Business Statistics-1	
Course Code: 302	**Author :** Dr. S. S. Tasak
Lesson No. : 5	**Vetter:** Dr. R. K. Mittal
INDEX NUMBER - I	

Structure

5.0 Learning Objectives

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

- Understand the concept of index number
- Explain the types and uses of index number
- Explain the different methods of preparation of index number
- Explain the problems in preparation of index numbers

5.1 Introduction

Index numbers are today one of the most widely used statistical indicators Generally used to indicate the state of the economy, index numbers are aptly called 'barometers of economic activity'. Index numbers are used in comparing production, sales or changes in exports or imports over a certain period of time. The role played by index numbers in Indian trade and industry is impossible to ignore. It is a very well-known fact that the wage contracts of workers in our country are tied to the cost of living index numbers.

5.1.1 Meaning

By definition, an index number is a statistical measure designed to show changes in a variable or a group or related variable with respect to time, geographic location or other characteristic such as income, profession etc. Index number is calculated as a ratio of the current value to a base value and expressed as a percentage. It must be clearly understood that the index number for the base year is always 100. An index number is commonly referred to as an index.

Index Number of wholesale prices

Base year 1981-82 (=100)

1992-93	Primary Articles	Manufactured Product	All Commodities
July	237.6	226.6	226.6
August	240.8	224.7	228.8
September	237.9	227.7	230.7
October	237.1	229.3	232.4
November	235.8	228.7	231.7
December	235.0	228.7	231.4
January	235.0	228.6	231.6
February	234.6	229.9	232.8
March	232.9	230.8	233.1
1993-94			
April	234.2	231.2	234.6
May	234.6	232.2	235.2
June	234.9	233.5	237.7

Source: CMIE, July 1993

Using these data, one can find out that the wholesale price for primary articles (comprising food articles, non-food articles and minerals) in April 1993 was 234.2 percent of the average wholesale price for primary articles in 1981-82. Similarly in June 1993, the wholesale price for all commodities was 237.7 times that of the wholesale price prevalent in the year 1981-82, on an average, for all commodities.

An index number is an average with a difference. An index number is used for purposes of comparison in cases where the series being compared could be expressed In different units. i.e., a manufactured products index (a part of the wholesale price index) is constructed using items like Diary products, Sugar, Edible Oils, Tea and Coffee etc. These items naturally are expressed in different units like sugar in kg, milk in liters etc. The index number is obtained as a result of an average of all these items which are expressed in different units. On the other hand, average is a single figure representing a group expressed in the same units.

Index numbers essentially capture the changes in the group of related variables over a period of time. For example, if the index of industrial production is 215.1 in 1992-93 (base year 1980-81) it means that the industrial production in that year was up by 2.15 times as compared to 1980-81. But it does not however mean that the net increaser in the index reflects an equivalent increase in industrial production in all sectors of the industry. Some sectors might have increased their production more than 2.156 times while other sectors may have increased their production only marginally.

Characteristics of Index Number-:

For a proper understanding of the index numbers, one should be clear about its characteristics. The following are the important characteristics of an index number:-

These are expressed in percentage: Index numbers are expressed in terms of percentages so as to show the extent of relative change.

1. *These are relative measure:* Index numbers are specialized averages used to show the relative change in group of related variables. The group of variables may relate to prices of certain commodity of volume of production of certain items. They compare changes taking place over time or between places. If the wholesale price index for the year 1990 is 140 as compared to 100 in 1988, then we conclude that the general price level has increased by 40% in two years.
2. *Index numbers are specialize averages:* Simple averages can be using those series which are expressed in the same units. However, index numbers are special type of averages which are used in comparing changes in series expressed in different units. In view of this they are also called specialized averages.
3. *Index numbers measure changes which are not directly measurable:* The index numbers are used for measuring the magnitude of changes in such phenomena which are not capable of direct measurement. For example, price level, cost of living and ups and downs in business activities are phenomena in which changes in directly measurable factor affecting price level, cost of living and business activities, indeed numbers help us to measure relative changes in corresponding

phenomenon which is otherwise not directly measurable.

5.1.2 Types and Uses of Index Numbers

Types of Index Numbers:

Index numbers are of different types. Important types of index numbers are discussed below:

1. Wholesale Price Index Numbers:

Wholesale price index numbers are constructed on the basis of the wholesale prices of certain important commodities. The commodities included in preparing these index numbers are mainly raw-materials and semi-finished goods. Only the most important and most price-sensitive and semi- finished goods which are bought and sold in the wholesale market are selected and weights are assigned in accordance with their relative importance.

The wholesale price index numbers are generally used to measure changes in the value of money. The main problem with these index numbers is that they include only the wholesale prices of raw materials and semi-finished goods and do not take into consideration the retail prices of goods and services generally consumed by the common man. Hence, the wholesale price index numbers do not reflect true and accurate changes in the value of money.

2. Retail Price Index Numbers:

These index numbers are prepared to measure the changes.in the value of money on the basis of the retail prices of final consumption goods. The main difficulty with this index number is that the retail price for the same goods and for continuous periods is not available. The retail prices represent larger and more frequent fluctuations as compared to the wholesale prices.

3. Cost-of-Living Index Numbers:

These index numbers are constructed with reference to the important goods and services which are consumed by common people. Since the number of these goods and services is very large, only representative items which form the consumption pattern of the people are included. These index numbers are used to measure changes in the cost of living of the general public.

4. Working Class Cost-of-Living Index Numbers:

The working class cost-of-living index numbers aim at measuring changes in the cost of living of workers. These index numbers are consumed on the basis of only those goods and services which are generally consumed by the working class. The prices of these goods and index numbers are of great importance to the workers because their wages are adjusted according to these indices.

5. Wage Index Numbers:

The purpose of these index numbers is to measure time to time changes in money wages. These index numbers, when compared with the working class cost-of-living index numbers, provide information regarding the changes in the real wages of the workers.

6. Industrial Index Numbers:

Industrial index numbers are constructed with an objective of measuring changes in the industrial production. The production data of various industries are included in preparing these index numbers.

The important uses of index numbers are described below:

1. They are economic barometer: Index numbers are mainly used in business and economics. Like barometers are used in physics ton measure atmospheric pressures, index numbers measure the level of business and economic activities and are, therefore, termed as ’economic barometers’ or ’barometers of economic activity’ For instance, price index numbers are those that relate to changes in the level of prices of commodity or a group of commodities over a period of time. Price index numbers are useful in studying price movements and determining their effect on economy. It is often useful to compare changes in general price level with changes in related series, such as ban deposits, bank loans, etc., for formulating economic policies.
2. The measure comparative changes: The important purpose of index numbers is to measure the relative change in a variable or a group of related variables in respect to time or place. The changes in the phenomena like price level, cost of living, etc., are not capable of being measured directly and are, therefore, measured with the help of index numbers. Indices of physical changes resulting over a period of time in production, sales, imports, etc., are extremely helpful in analyzing the movements in these characteristics over time.
3. *They help in forecasting:* Many governmental and private agencies are engaged in computation of index numbers for purpose of forecasting business and economic condition. For example, index number of industrial and agriculture production not only reflect the trend but can also help in forecasting future production Similarly, index numbers of unemployment in a country not only reflect the trends in the phenomenon but are useful in determining factors leading to unemployment. The analysis of such trend and factors in unemployment activity help in framing a suitable employment policy.
4. *They measure the purchasing power of money*: Consumer price index numbers are useful in finding the intrinsic worth of money as they are used for adjusting the original data related to wages for price changes. In other words, index numbers are helpful for transforming nominal wages into real wages. Based on this aspect, the Government of India of the states use consumer price index numbers for

determining the amount of additional wages or dear ness allowances to be given to their employees to compensate for changes in the price level or cost of living. Thus, most of the government now has index linked salary structures and additional dearness allowance is granted to employees for a point rise in consumer price index.

5. *They measure the real gross national product (G.N.P.):* Index numbers are also used for determining the real gross national product (G.N.P) or income calculated at current prices. Real G.N.P. is determined by price index of the current year., i.e.,

$$\text{Real G.N.P.} = \frac{\text{G.N.P. at current price}}{\text{Current price index}} * 100$$

5.1.3 Method of constructing index numbers

A number of formulae have been developed for constructing index numbers which may be grouped into the following categories-

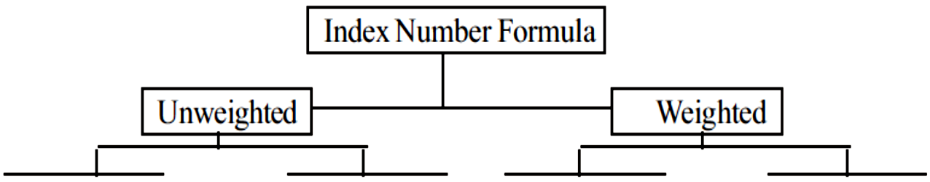
1. Simple or unweighted index numbers
2. Weighted index numbers

Various index formulae in each of the above to categories may be further classified as-

(a) Simple aggregative method

(b) Method of simple average relatives

The following chart may be used define the above classification-



Simple aggregative, Simple average of relatives, Simple aggregative, Simple aggregatives relatives

Notations :

For proper understanding of various index number formulas, the understanding of the following notations is necessary:

Po(j) : Price of jth commodity in the base year, (j=1,2,......N)

Pn(j) : Price of jth commodity in the current year, (j=1,2. ... N)

qo (j) : Quantity of jth commodity consumed or purchased in the base year:,(j1,2,. ....N)

qn(j) : Quantity of jth commodity consumed or purchased in the current year., (1,2. N)

w(j) : Weight assigned to jth commodity according to its relative importance, (j=1,2. N)

Pon : Price index for the current year (n) with respect to the base year (O) Pno : Price index for the base year (O) with respect to the current year (n) Qon : Quantity index for the current year (n) with respect to the base year (O) Qno : Quantity index for the base year (O) with respect to the current year (n) Von : Value index for the current year (n) with respect to the bas year (O)

Vno : Value index for the base year (o) with respect to the current year (n).

Price Relatives, Quantity Relatives and Value Relatives

<u>Price Relatives :</u>

Price relative is one of the simplest example of an index number. It is defined as the ratio of the price of a single commodity in the current year (n) to its price in the base year (o). Therefore, the price relative of period (n) with respect to period (o) is

$$\text{Price Relative} = Pon = \frac{Pn}{Po} \quad \ldots\ldots\ldots\ldots(1)$$

It is customary to express price relative as percentage by multiplying it by 100.

Thus, the price relative in (1) expressed in percentage becomes

$$\text{Oruce Relative} = Pon = \frac{Pn}{Po} X\ 100 \quad \ldots\ldots\ldots\ldots(2)$$

Here it is important to note that the price relative for a given period with respect to same period is always 1 or 100 in percentage terms. In particular, we can say that the price relative of the base year is always 100. In defining the price relatives, the prices are assumed constant at one time point. However, if they very over a period, an appropriate average of the prices over the given duration is used for the purpose.

Illustration 1. Let the prices of certain item in 1985 and 1990 were Rs. 110 respectively Then taking the year 1985 as base, m the price relative for the year 1990 will be -

Price Relative for the year 1990=Pon= $\frac{Pn}{Po}$ X 100 = $\frac{110}{70}$ X 100 = 157.14%

As Pn= the price in the current year, i.e., in 1990=110
Rs. and Po=the price in the base year 1985 = 70 Rs.

Illustration 2. The Price of two commodities in the year 1990 and 1991 are shown in the table. The price relatives of the two commodities are also given assuming 1990 as the base year-

Table Showing Price Relative

Price			Price Relative
1990 (Po)		1991(Pn)	Pon=Pn/PoX100 118.18%
A	55	65	117.11%
B	76	89	

Quantity Relatives

Quantity relative are another type of index numbers used for measuring changes in quantum or volumes of a commodity, such as quantity of production, consumption. exports, imports, sales, etc. Thus, in this case the commodity is used in a more general sense. It may mean volume, of exports, imports, sales, production, and number of passengers travelling by railways and so on. As in the case of prices, quantities too are assi, ed constant for any period, otherwise, an appropriate average is used for the purpose to mae this assumption valid. Now, the quantity relative is defined as the ratio of the quantity of a single commodity in the current year (n) to its quantity in the base period (o), thus,

Quantity Relative=Qon=qn/qo. (3)

Expressing this ratio in percentages, we get,

Quantity Relative=Qon=qn/qo X 100 ....(4)

Quantity relatives too have the same properties as that pertaining to price relatives.

Illustration 3. Let us consider the data on production of wheat of a country in millions of tons for three years-

Years	1989	1990	1991
Production of Wheat (Millions of tons)	1090	988	1306

Based on the above data, the quantity relatives for the year 1990and 1991 taking 1989 as the base are shown in the following table.

Years	1989	1990	1991
Production of wheat	1090(qo)	988(ql)	1306 (q2)

Quantity Relative $Q00=\frac{1090}{1090}X100$ $\frac{988}{1090}X100Q02$ $\frac{1306}{1090}X100$

=100% =90.64% =119.82%

Value Relatives

A value relative is yet another type of index number. It is used in a situation when one is willing to compare changes in the monetary value of consumption, sale export, and import of commodity on two or more points of time. If p and q are the price and quantity of a commodity produced, consumed or sold, during a period, then v = pq gives the total money value of the transaction. Her a also p and q are assumed constants at a time point, otherwise a suitable average of prices and quantities over the given period is taken to make this assumption valid. In usual notations, then we define.

Vo=poqo : total money value during the base period.

vn = pnqn : total money value during the current period.

Therefore, the value relative of the current year is the ratio of vn to vo, i.e.

$$\text{Value Relative} = \underline{V0n} = \frac{vn}{vo} \quad \text{.. (5)}$$

When expressed in percentage terms, one gets

$$\text{Value Relative} = V0n = \frac{vn}{Vo} \times 100 \quad \text{………………….(6)}$$

Solution :

	Price Per unit (Rs.)		Quantity consumed		Values in (Rs.)		Value Relatives $Qon=\frac{vn}{vo}X\ 100$ In (unit)
Commodity	1980	1985	1980	1985	1980	1985	
	Po	Pn	qo	qn	vo=poqo	vn=pnqn	
A	7	10	40	52	280	520	$\frac{520}{280}$ x 100
							=185.71%
B	9	12	75	80	675	960	$\frac{960}{675}$ X 100
							=142.22%

Properties of Relatives

If pa, pb, pc,.........be the respective prices of the commodity in periods a, b,c., then, with usual notations, the price relative have the following properties :-

1. Identity property :

If pa, pb, pc,.........be the respective prices of the commodity in periods a, b,c. , then, with usual notations, the price relative have the following properties :-

1. Identity property :

According to this property, the price relative for a given period with respect to the same period is 1 or 100% That is -

	$Paa= \frac{pa}{pa} =1$	or	$Paa = \frac{pa}{pa} X\ 100 = 100\%$
or	$Pbb = \frac{pb}{pb} =1$	or	$Pbb = \frac{pb}{pb} X\ 100 = 100\%$

1. Time Reversal property

This property states that if the current period and the base period are interchanged, then the product of the corresponding price relatives is unity. In other words, the corresponding price relatives are reciprocal of each other. Symbolic.

Pab = pb and Pba = pa

pa pb

: $Pab \times Pba = \frac{Pb}{Pa} \times \frac{Pa}{pb} = 1$

or $Pab = 1/Pab$ and $Pba = 1/Pab$

2. Cyclical or Circular Property

According to this property, if the periods a., b, and c are in circular order then the product of the three relative defined with respect to the preceding period as base period is unity. In Symbols :

$$Pab. Pbc. Pca = 1$$

as $Pab = \frac{pb}{pa}$, $Pbc = \frac{pc}{pb}$, $Pca = \frac{pa}{pc}$

The property can be extended to any number of periods which are in circular order, For example, for four periods a, b, c, and d which are in circular we have

$$Pab. Pbc. Pcd. Pda = 1$$

3. Modified Cyclical or circular property

Using the concepts in properties 2 and 3, we have the following modified circular property of the price relatives.

$$Pab. Pbc = Pac$$

In case of four periods,

$$Pab, Pbc. Pcd = Pad$$

Illustration 5. Let us consider the following data related to prices of a commodity of three years-

Year	1980	1985	1990
Price (Rs.)	54	67	90

With the help of given data, let us check the properties of the pr relaives as discussed above. Symbolic, the given can be put as-

a = 1980	b = 1985,	c = 1990	
And	pa = 54,	pb = 67,	pc = 90

Indentity property :

Here,

$$Paa = \frac{pa}{pa} X\ 100 = \frac{54}{54} X\ 100 = 100\%$$

$$Pbb = \frac{pb}{pb} X\ 100 = \frac{67}{67} X\ 100 = 100\%$$

$$Pcc = \frac{pc}{pc} X\ 100 = \frac{90}{90} X\ 100 = 100\%$$

Times Reversal Property :

$$Pab = \frac{pb}{pa} = \frac{67}{54}; \qquad Pba = \frac{pa}{pb} = \frac{54}{67}$$

Obviously Pab = 1/Pba.

Similarly, we can observe that Pbc = 1/Pcb and Pac = 1/Pca

Cyclical or circular property :

Since,

$$Pab = \frac{pb}{pa} = \frac{67}{54},\ Pbc = \frac{90}{67} \qquad Pca = \frac{54}{90}$$

$$Pab\ X\ Pbc\ X\ Pca = \frac{67}{54} X \frac{90}{67} X \frac{54}{90} = 1.00$$

Modified Circular Property :

Let us consider

$$Pac\ X\ Pbc = \frac{67}{54} X \frac{90}{67} = \frac{90}{54}$$

Also

$$Pac = \frac{pc}{pa} = \frac{90}{54}$$

Pab X Pbc = Pac

5.1.4 Simple or Unweighted Index Number

Simple aggregate Method

This is the simplest method of computing index numbers. According to this method the total of the current year prices for various commodities is expressed as a percentage of the total of the base year prices for these commodities. In Symbols,

$$\text{Simple aggregate price index Pon} = \frac{\Sigma Pn}{\Sigma Po} X\ 100$$

Here Pon = Current year index number

Pn = the total of commodity prices in the current year

Po = the total of commodity prices in the base year.

Example 1 : Using simple Aggregate method obtain index numbers for 1990 taking 1988 as the base year from the following data -

Commodities		A	B	C	D	E
Price (Rs.)	1988	100	80	160	220	40
	1990	140	120	180	240	40

Solution : Computing Index number (Simple Aggregate Method)

		Price (Rs.)
Commodities	1988 (Po)	1990 (Pn)
A	100	140
B	80	120
C	160	180
D	220	240
E	40	40
Total	Σ Po = 600	ΣPn = 720

$$\text{Price Index for 1990} = \frac{\text{Pon}}{\Sigma \text{ Po}} = \Sigma \text{ Pn X } 100 = \frac{720 \text{ X } 100}{600} = 120$$

Which shows a net increase of 20% in the price of commodities in the year 1990 as compared to 1988.

Example 2 : From the following data construct price index numbers for the years 1985 and 1990 by simple aggregate method taking 1980 as the base year -

Price (in Rs.)

Commodities	Unit	1980	1985	1990
Wheat	Quinital	200	250	275
Rice	Quintal	300	350	450
Arhar	Quintal	600	700	750

Milk	Litre	6	7	8
Clothing	Metre	30	35	40

Solution :

Construction of Price Indices (Simple Aggregate Method)

Construction of Price Indices (Simple Aggregate Method)

		Price (In Rs.)	
Commodities	1990 (P0)	1985 (P1)	1990 (P2)
Wheat	200	250	275
Rice	300	350	450
Arhar	600	700	750
Milk	6	7	8
Clothing	30	35	40
Total	$\Sigma Po = 1136$	$\Sigma P1 = 1342$	$\Sigma P2=1523$

: Price Index for 1985 = P01 = $\frac{\Sigma P1}{\Sigma P0} \times 100 = \frac{1342}{1136} \times 100 =$ 118.13

Price Index for 1990 = $\frac{\Sigma P_2}{\Sigma P_0} \times 100 = \frac{1523}{1136} \times 100 =$ 134.07

Price Index for various years thus becomes -

Year	:	1980	1985	1990
Index number (base 1980)	:	100	118.13	134.07

Example 3 : Prices of and article for six years are given below :

Year:	1980	1981	1982	1983	1984	1985
Prices (in Rs.) :	10	14	16	20	22	26

Obtain price index numbers (i) assuming 1980 as the base year, (ii) assuming the average price of the six years as base.

Solution : The computation of the price index numbers in the two cases is shown in the following table - Construction of Price Index Num

Year	Price	Index Numbers	Index Numbers

		(base 1980) (pn/po) X 100	(Average Price 18
			as 100)
1980		100	10 X 100 = 55.5 18
1981	14	14 X 100 = 140 10	14 X 100 = 77.77
1982	16	16 X 100 = 160 10	16 X 100 = 88.88 18
1983	20	20 X 100 = 200	20 X 100 = 111.11
1984	22	10 22 X 100 = 220	18 22 X 100 = 122.22
1985	26	18 26 X 100 = 260	18 26 X 100 = 144.44
		10	18

$$\text{Average price of the six years} = \frac{10 + 14 + 16 + 20 + 22 + 26}{6}$$

$$= \frac{108}{6} = 18 \text{ Rs.}$$

Limitations of the simple Aggregate Method

Although the method is simple, it has the following two limitations -

1. The above formula does not consider the relative importance of various commodities involved.
2. In this formula, the prices of various commodities are generally given in different units, e.g., wheat may be quoted in Rs. per quintal ; milk, petrol in Rs. per litre ; cloth in Rs. per meter and so on. Thus the particular units used in the price quotations may affect the value of the index number.

Simple Average of price relatives Method

In this method, we first obtain price relatives for all commodities included in the index numbers. As usual, the price relative off the current year expressed as a percentage of the price of the base year is pn X 100. Now these commodity price relatives may be averaged by using averages like arithmetic mean, median mode or geometric mean. Howver, if we use arithmetic mean, the formula for computing index number becomes –

$$Pon = \frac{\sum \frac{pn}{Po} \times 100}{N}$$

$$Pon = \frac{\sum \frac{pn}{Po} \times 100}{N}$$

Here, $\sum \frac{Pn}{Po}$ = the sum of commodity price relatives.

N = The number of commodities.

Similarly, the geometric mean of the relatives can be used to find the index number of the current year as-------

$$Pn = \text{Anti log} \left\{ \frac{\sum \log \frac{Pn}{Pn} \times 100}{N} \right\} = \text{Anti log} \left\{ \frac{\sum \log R}{N} \right\} \text{; where } R = \frac{Pn}{Po} \times 100$$

Example 4 : Construct index number for 1990 taking 1988 as the base year from the following data by using average for price method.

Commodities		A	B	C	D	E
Price in Rs.	1998	100	80	160	220	40
	1990	140	120	180	240	40

<u>Construction of Price Indices (Simple average of price relation)</u>

Commodities Price in 1998 Price in 1990 Price Relatives

A	100	140	140X100=140.00
B	80	120	120X120=150.00
C	160 (Base) Po	180 (Current Year) Pn	180x100=112.50 (Pn/Pnx 100)
D	220	240	240x100=109.10
E	40	40	40x100=100.00
N=5			ΣPn x 100 = 611.60

Price Index No. for 1990 = (Pon) = (Pn x 100) = 611.60 = 122.32
Po
N

Example 5 : Use the following information to construct index number for 1990 taking the price of 1985 as base. Use (i) simple aggregate method and (i) simple average of relative method in the construction.

Commodity	A	B	C	D	E
Price in Rs.	1985	12	25	10	6
	1990	15	20	15	15
Solution :					

Construction of Index numbers :

Commodity Price in 1985 (base year) Price in 1990 (Current Year)

	Price Po	Price Relatives	Price Pn
A	12	100	15
B	25	100	20
C	10	100	12
D	5	100	10
E	6	100	10

Price Relatives Pnx100 15
x 100 = 125.00
12
20 x 100 = 80.00
25
12 x 100 = 120.00
10
10 x 100 = 200.00
5
15 x 100 = 250.00
6

N=5 ∑PO=58 SPn = 72 Pn x 100 = 775
Price index for 1990 SPO
Σ Pn 72

(i) By aggregate method = Pon = Σ P0 X 100= 58 X 100 = 124.14

(ii) By simple average of price relatives Pon = $\Sigma \left(\frac{\frac{Pn}{Po} \times 100}{N} \right) \frac{775}{5} = 155.50$

Example 6. Using arithmetic mean, median and geometric mean, construct index numbers by the simple average of relative method from the following data for 1990 and 1991 with 1989 as the base year.

Price (in Rs. per unit)

Commodity	1989	1990	1991
A	100	120	150
B	40	45	60
C	30	35	45
D	10	12	15
E	20	22	23

Solution :

Computation of price Indices [(Using Mean, Median and G.M. (base 1989)] Articles Prices (Rs.) Price Relatives for 1990 Price Relative for 1991

	P0	P1	P2	R1=P1/P0 X 100	log 21	R2=P2/PoX100	log R2
				120		150	
A	100	120	150	120 X 100 = 120.0	2.0792	X 100 = 150.00	2.1761
				100		100	
B	40	45	60	45/50 X 100 = 112.5	2.0511	60/40X100=150.0	2.1761
C	30	35	45	35/30 X 100 = 116.7	2.0671	45/30X100=150.0	2.1761
D	10	12	15	12/10 X 100 = 120.0	2.0792	15/10X100=150.0	2.1761
E	20	22	23	22/20 X100	=110.0	2.041423/20X100	= 2.060

N=5	ΣR1 = 597.2	ΣLog R1 = 10.318	ΣR2=715	Σlog R2= 10.7651

(i) Price Index by using arithmetic mean of the relatives.

Price Index for 1990 = P0 = $\frac{\Sigma P1/P0}{N}$ x 100 = $\frac{\Sigma R1}{N}$ = 579.2 = 115.80

Price Index for 1991 = P0 n = $\frac{\Sigma P2/P0}{N}$ x 100 = $\frac{\Sigma R2}{N}$ = 715.0 = 143.0

(ii) Price Index by median of the relatives.

Arranging the price relative in ascending order, and selecting the size of $\frac{N+1th}{2}$ = 3rd term, we have

Price Index for 1990 = 116.70

Price Index for 1991 = 150.00

(iii) Price Index by using G.M. of the relatives.

Price Index for 1990 = anti log ($\frac{\Sigma \log R1}{N}$) = anti log ($\frac{10.3180}{5}$)

anti log [2.0636] = 115.8

Price Index for 1991 = anti **log**(ΣR 2) = anti log (10.7651) N 5

anti log [2.1530] = 142.2

Thus, price indices for various years can be summarised as -

Used Average		Price indices (Years)	
	1989	1990	1991
A.M.	100	115.8	143.0
Median	100	116.7	150.0
G.M.	100	115.8	142.2

Example 7. Taking of I year as base, construct the index numbers for II and III years from the following data. Use the simple average of relatives method.

Year Articles (Rate per Ruppes)

	A	B	C
I	4 kg	2 kg	1 kg
II	2.5 kg	1.6 kg	1 kg
III	2.0 kg	1.25 kg	0.8 kg

Solution : In this example, prices are given in quantity per rupee'. Thus, before computing price relatives, these are to be converted into 'rupees per unit of quantity'. Considering the unit of the quantity as 'quintal' the prices in 'rupees per quintal' are given below - Year Price (in Rupees per quintal)

	A	B	C
I	25	50.0	100
II	40	62.6	100
III	50	80.0	125

Now, the price index numbers for II and III years (I years as base) can be constructed as usual. The procddure is clarified in the following table-

Construction of price Relatives

Price Relatives for II Price Relatives for III

Price (Rs.) year (1 year as base) Year (I year as base)

Article

	I	II	II	R1=P1/Po X 100	R2=P1/P0 X 100
A	25	40.0	50.0	160.0	200.0
B	50	62.5	80.0	125.0	160.0
C	100	100.0	125.0	100.0	125.0

Total N =3 S R1 = 385.0 485.0

$$\text{Price Index for II year} = P01 = \frac{\Sigma P1}{N}/P0X100 = \frac{\Sigma R1}{N} = \frac{385.0}{3} = 128.33$$

$$\text{Price Index for III year} = P02 = \frac{\Sigma\ P2/Po\ X\ 100}{N} = \frac{\Sigma\ R2}{N} = \frac{485.0}{3} = 161.67$$

Merits and Demerits of Simple Averages of Relative Method

Merits

1. The Index number is not influenced by extreme items. All items are given equal weightage.
2. The method provides an index which is not affected by the particular units used in price quotations.

Limitations

1. In this method, we face with the problem of selecting a suitable average.
2. All commodity price relatives are given equal weightage which may not be true always.

5.2 Problems in the Construction of Index Number

Before constructing index numbers, a careful study of the following related problems is made:

1. The purpose of the index number: Index numbers are constructed for serving specific purposes; therefore, it is important to know what kind of changes we are trying to measure and how we intend to use them. Obviously, a clear definition of the purpose and objective is the first major problem in the construction of index numbers. For example, if price index is to be constructed for measuring the cost of living of middle class families in a region, care must be taken to include items which are consumed by these families. Similarly, for measuring prices and not wholesale prices of these items.

2. Selection of items: Having defined the purpose of index numbers, the next problem relates to the Selection of items, In this regard, a decision about the number of items included, the more representative shall be the index but at the same time cost and time involved in the construction of index number will increase. Therefore, the number of times selected should neither be too small nor too large. Secondly, one should ensure that the items selected represent the tastes, habits and customs of the people for whom the index is being constructed. For instance, while computing cost of living index for middle class families, gold, car, etc. will not be relevant items. Thus, only relevant standardized items, which are easy to define and describe, should be included in the construction of index numbers so that they reflect the change that we wish to measure.

3. Data for index numbers: The data used in index numbers are usually concerned with the prices and quantities consumed of the selected items for different points of time. As such, we always face the problems of selecting a reliable source of data. Thus, the data should be collected from standard trade journals, official publication, chamber of commerce and other government agencies. Data are also collected through field studies or sample surveys. Here the samples selected should be representative of the class to which they belong and then only the resulting data is expected to be reliable, accurate and homogeneous. For uniformity, it is often desirable to group the items into homogeneous groups of subgroups. For instance, in measuring price changes, domestic items may be grouped into cereals, mill. Edible oils, clothing, electricity and fuel, etc. similarly, items with elastic demand. Type of price quotations is another consideration while collecting data. A wholesale price index needs wholesale price quotations while retail price quotations will be desirable will be desirable in the construction of cost of living index number.

4. Choice of base period: The period with which the comparison of the relative changes in the level of a phenomenon is made is termed as 'base period of 'reference period'. The index for the base period is always taken as 100. For reliable and precise comparisons of the relative changes, a base year should be a sufficiently 'normal' year. It should be period free from all abnormalities like economic boom or depression, labor strikes, wars, earthquakes, etc. In other words, the base period free from all abnormalities, etc. In other words, the base period should be more or less stable and free from unusual ups and downs. It is also desirable that the base period should not be too distant from the given period with which relative changes are measured. In case the base period is too distant, it is desirable to shift it, For example, it seems undesirable to compare prices of commodities must have changed since then. Notable technological developments, rising income of the people, changing patterns of consumption, quality of goods, changes in habits and tastes of the people are some factors which compel the shifting of the base period. The base period may be of two types:

- Fixed base period.
- Chain base period.

{i} Fixed base period: If the base period or reference period is kept fixed for all current periods of comparison, it is called the fixed base period. For example, the year 1951, being the first year of planning process, may be taken as the base period for studying relative planning development in the current years.
{ii} Chain base period: In chain method, the change in the level of the phenomenon for any given period is compared with the level of the phenomenon in the preceding period and not to the base period.

5. Choice of an average: We observed that index numbers are special type of averages. An, the choice of a suitable average is also important in the construction of index numbers. Arithmetic mean, median and geometric mean are the commonly used averages in index numbers. Out of the three averages, arithmetic mean and a median are comparatively easier to calculate. However, median completely ignores the extreme observations while the arithmetic mean is unduly affected by such observation. The geometric mean is the most suitable average in the construction of index numbers in view of the following properties:

{i} Geometric mean gives more importance to smaller items and less to larger items and is, therefore, least affected by the values of the extreme items.

{ii} Geometric mean gives equal weights to equal ration of changes.

In spite of theoretical justification for its suitability, geometric mean is not a common average in the construction of index numbers. It is view of its difficult computational process for simplicity in calculations. Arithmetic mean is used instead, however, geometric mean is recommended for greater accuracy and precision.

6. Selection of weights: Unweighted index numbers give equal importance to all commodities. However, all items or commodities included in the construction of an index number are not of equal importance. For example, in the construction of cost of living index, sugar cannot be given the same importance as the cereals. In order to allow each commodity to have a reasonable influence on the index, we make use of weighted index numbers which give appropriate weight to different commodities according to their importance. The selection of appropriate weight is again a difficult task. The methods of assigning weight are:

{a} Implicit (or arbitrary) {b} Explicit (or actual).

In implicit weighting, a commodity or its variety is included a number of times according to its importance. On the other hand, in explicit weighting, some actual criteria are used for assigning weights to different commodities included in the index number.

Generally, the weights for various commodities are decided according to {i} Value or produced, {ii} Value or quantity consumed, {iii} Value or quantity sold). When the quantity (value) is the basis of weight, we call it quantity as weights. On the other hand, in the method of averaging price relatives, values are used as weights

7. Selection of suitable formula: Selection of a suitable formula for construction of any index number also poses some problems. There are various formulae for calculating index numbers such as the aggregate method or the average of relatives method in simple aggregate method the price of each commodity is given in usual units and this leads to the dominance of a particular quantity in the index. The difficulty is removing by considering the average of relatives methods. But in this method we assume that each item is purchased for an equal amount of money in the base year i.e. the value of all the items in the base year is the same. This leads to the concepts of weighted in ex formula where items are waited according to their relative importance in this regard a separate section is devoted for method of constructing index No. there suitability in a particular situation.

5.3 Check Your Progress

(1) Index number is calculated as a ratio of the …………………… to a base value and expressed as a percentage.

(2) Industrial index numbers are constructed with an objective of measuring changes in the………………………………

(3) Simple aggregate method can be applied only when the prices of all the commodities have been expressed in the…………………….

(4) The purpose of wage index numbers is to measure time to time………………. in money wages.

(5) The working class cost-of-living index numbers aim at measuring…………………… in the cost of living of workers.

5.4 Summary

An index number is a statistical measure designed to show changes in a variable or a group or related variable with respect to time, geographic location or other characteristic such as income, profession etc. There are different types of index numbers like Wholesale Price Index Numbers, Retail price index numbers, Cost-of-Living Index Numbers, Working Class Cost-of-Living Index Numbers, Wage Index Numbers and Industrial index numbers. There are different methods of constructing index numbers which are categorized into two parts. First is simple or unweighted index number which is further categorized into simple aggregate method and simple average of price relative method and second is weighted index numbers which are further categorized into weighted aggregate index and Weight Average of Relative. Before constructing index numbers, a careful study of the following related problems is made: the purpose of the index number, Selection of items, Data for index numbers, Choice of base period, Choice of an average, Selection of weights, and Selection of suitable formula.

5.5 Keywords

Wholesale price index: This number is constructed on the basis of the wholesale prices of certain important commodities.

Retail price index: These index numbers are prepared to measure the changes in the value of money on the basis of the retail prices of final consumption goods.

Cost-of-Living Index Numbers: These index numbers are constructed with reference to the important goods and services which are consumed by common people.

Working Class Cost-of-Living Index Numbers: The working class cost-of-living index numbers aim at measuring changes in the cost of living of workers.

Wage Index Numbers: The purpose of these index numbers is to measure time to time changes in money wages.

Simple Aggregate Method: In this method, sum of current year' prices is divided by the sum of base year's prices and the quotient is multiplied by 100.

5.6 Self- Assessment Test

Q1. Explain the term index numbers.

Q2. Explain the methods of construction of index numbers.

Q3. Write a short note on the following:

(a) Simple aggregate method

(b) Simple average of price relatives method

Q4. Prepare price index numbers for three years taking the average price as base using simple average of price relative method:

Price per Quintal (Rs.)

Year	Wheat	Cotton	Oil
1995	100	25	30
1996	99	20	25
1997	99	15	30

(Ans. 110.52, 95.97, 93.52**)**

Q5. Construct the price index number for three years taking the average price as base by using simple average of price relative method:

Year	Wheat	Rice	Sugar
I	2 kgm	1 kgm	0.400 kgm
II	1.6 kgm	0.800 kgm	0.400 kgm
III	1 kgm	0.750 kgm	0.250 kgm

(**Ans**. 79.2, 92.1, 128.7)

5.7 Answers to Check Your Progress

(1) Current value

(2) Industrial production

(3) Same unit

(4) Changes

(5) Changes

5.8 References/Suggested Readings

1. Gupta, S. P.: Statistical Methods, Sultan Chand and Sons, New Delhi.
2. Levin, R. I. and David, S. R.: Statistics for Management, Prentice Hall, New Delhi.
3. Gupta, C. B.: Introduction to Statistical Methods.
4. Hooda, R. P.: Statistics for Business and Economics, Macmillan, New Delhi.

Subject: Business Statistics-1	
Course Code: 302	**Author :** Dr. S. S. Tasak
Lesson No. : 6	**Vetter:** Dr. R. K. Mittal
INDEX NUMBER-II	

Structure

6.0 Learning Objectives

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

- Understand the types of weighted index number
- Explain the tests of adequacy of index number
- Explain the limitations of index number

6.1 Introduction

As observed, simple or unweighted index number assign equal importance to all the commodities or items included in the index. However, various commodities included are not of equal importance. To overcome this disadvantage of the simple index numbers, we weight the price of each commodity by a suitable factor. This factor is generally taken as the quantity or volume of the commodity sold or consumed during the base year, the current year or some typical year (taken as an average over a number of years). In this way, the importance of the commodities as also reflected in the index number.

6.1.1 Types of weighted index numbers

As indicated earlier, weighted index number too can be classified as:

1. Weighted aggregate index
2. Weight Average of Relatives

6.1.2 Weighted Aggregate Index Numbers

In this method, appropriate weights are assigned to various commodities which reflect their relative importance in the group. A reasonable assumption is to consider the quantities consumed or produced as weights. If W is the weight attached to commodity then a general weighted price index can be formulated as under :-

$$\textbf{Weighted aggregate price Index} = P_{0n} = \frac{\sum P_a W}{\sum P_0 W} \times 100 \qquad ....(1)$$

Where symbols have their usual meanings.

By using different types of weight in (9), a number of formulas have been developed for the construction of index numbers, these formulas are -

1. Laspeyers' index or base year method.
2. Pasche's index or given year method.
3. Marshall - Edgeworth's index number.
4. Walsh's index number.

5. Bowley's index number.
6. Fisher's ideal index number.

Laspeyres' Index Number

Taking the quantity of the base year i.e. q_0 as weight in formula (1), we get Laspeyre's price index number as –

$$\text{Laspeyres' Price Index} = P_{0n} = \frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times 100 \qquad ....(2)$$

Paasche's Index Number

According to this formula, the quantity of the given year, i.e. q_n is taken as weight. So using q_n for W in (9), we gent Paache's index as –

$$\text{Paasche's Price Index} = P_{0n} = \frac{\Sigma P_a q_n}{\Sigma P_0 q_n} \times 100 \qquad ....(3)$$

Marshall-Exgeworth's Index Number

In this formula, the average of the base year and given year quantity is taken as weight. Thus, on putting w = $(p_0+q_n)/2$ in (1), we gets

$$\text{Marshall-Edgeworth's Price Index} = P_{0n} = \frac{\Sigma P_n\left(\frac{q_0+q_n}{2}\right)}{\Sigma P_0\left(\frac{q_0+q_n}{2}\right)} \times 100$$

$$= P_{0n} = \frac{\Sigma P_n(q_0+q_n)}{\Sigma p_0(q_0+q_n)} \times 100 \qquad ....(4)$$

Walsh's Index Number

Taking as weights the geometric mean √ $[q_0 q_1]$ of the base and given year quantities, i.e., putting W = $\sqrt{[q_0 q_1]}$ in formula (1), one gets

$$\text{Walsh's Price Index} = P_{0n} = \frac{\Sigma P_n[\sqrt{(q_0+q_n)}]}{\Sigma P_0[\sqrt{(q_0+q_n)}]} \times 100 \quad \text{....(5)}$$

Bowley's Index Number

Taking the grithmetic mean of Laspeyres' and Paache's index numbers in (2) and (3) respectively, one gets

$$\text{Bowley's Price Index} = P_{0n} = \frac{1}{2}\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} + \frac{\Sigma p_n q_0}{\Sigma p_0 q_n}\right] \times 100 \quad \text{....(6)}$$

Fisher's Ideal Index Number

We define Fisher's ideal price index as the geometric mean of Laspyre's and Paasche's index numbers in (10) and (11). Therefore,

$$\text{Fisher's Price Index} = P_{0n} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} + \frac{\Sigma p_n q_0}{\Sigma p_0 q_n}\right]} \times 100 \quad \text{....(7)}$$

As will be discussed later, Fisher's ideal index satisfies both the time reversal test and factor reversal test which provides this index number a theoretical advantage over other formulas. In view of these properties, it is called an index formula.

Use of Laspeyres's and Paasche's Index Numbers

The main advantage of Laspeyre's index formula is that the weight q_0 for the base year remains the same throughout. Thus, while constructing the index number, only the changes in prices are experienced. On the other hand, for Paasche's index, weights q_n are also obtained for each given year. In case the base year is the typically selected normal year, then the use of base year quantities provides more stability to the index number. Obviously, the use of Laspeyres' index number is advantageous when the base year is stable and normal. On the other hand, if the conditions have been changing fast over the years, then considering current year's quantities (q_n) as weights, represents a more realistic picture. As such, Paasche's index is more suitable in this situation.

Example 8 : Construct an index number for 1990 (base year = 1980) from the following data using weighted aggregative index number.

Commodity	Weight	Prices (Rs.)	
		Base year (1980)	Current year (1990)
A	30	4.25	5.20
B	40	2.90	3.75
C	15	2.15	1.95
D	15	8.85	8.10

Solution: **Construction of Index**

Commodity	Weight	Prices (Rs.)		P_0W	P_nW
		Base year (1980) P_0	Current year (1990) P_n		
A	30	4.25	5.20	127.50	156.00
B	40	2.90	3.75	118.00	150.00
C	15	2.15	1.95	32.25	29.25
D	15	8.85	8.10	132.75	121.50
Total	100			$\sum P_0W$ =410.5	$\sum P_nW$ 456.73

$$\textbf{Weighted Index Number for 1990} = P_{0n} = \frac{\sum P_nW}{\sum P_0W} \times 100 = \frac{456.75}{410.50} \times 100 = \mathbf{111.27}$$

Example 9 : Construct price index numbers for one year 1990 (base year - 1985) from the following data by (i) Laspeyres' Method (ii) Paashe's method (iii) Marshall-Edgeworth's method (iv) Fisher'smethod.

Commodity	Base year (1985)		Current year (1990)	
	Price	Quantity	Price	Quantity

A	10	30	12	50
B	8	15	10	25
C	6	20	6	30
D	4	10	6	20

Construction of Price Index Numbers

Commodity	Base year		Current year					
	Price P_0	Quantity q_n	Price P_0	Quantity q_n	P_0q_n	P_nq_0	P_0q_n	P_nq_0
A	10	30	12	50	300	500	360	600
B	8	15	10	25	120	200	150	250
C	6	20	6	30	120	180	120	180
D	4	10	6	20	40	80	60	120
Total					Σp_0q_0	Σp_0q_n	Σp_nq_0	Σp_nq_0
					=580	=960	=690	=1150

(i) Laspeyre's Method - Using formula (2),

$$\textbf{Price Index} = P_{0n} = \frac{\Sigma p_nq_0}{\Sigma p_0q_0} \times 100 \qquad = \frac{690}{580} \times 100 = \mathbf{118.96}$$

(ii) Paasche's Method - Using formula (3),

$$\textbf{Price Index} = P_{0n} = \frac{\Sigma p_nq_n}{\Sigma p_0q_n} \times 100 \qquad = \frac{1150}{960} \times 100 = \mathbf{119.79}$$

(iii) Bowley's Method - Using formula (6),

$$\textbf{Price Index} = P_{0n} = \frac{1}{2}\left[\frac{\Sigma p_nq_0}{\Sigma p_0q_0} + \frac{\Sigma p_nq_n}{\Sigma p_0q_n}\right] \times 100$$

$$= \frac{1}{2}[(118.96 + 119.79)] \text{ X } 100 = 119.37.$$

(iv) **Marshall Edgewo -** Method Using formula (4),

$$\textbf{Price Index} = P_{0n} = \frac{\Sigma p_n(q_0+q_n)}{\Sigma p_0(q_0+q_n)} \times 100 = \frac{\Sigma p_n q_0 + \Sigma p_n q_n}{\Sigma p_0 q_0 + \Sigma p_0 q_n} \times 100$$

$$= \frac{690+1150}{580+960} \times 100 \quad = \mathbf{119.48}$$

(v) **Fisher's Method -** Using formula (7),

$$\textbf{Price Index} = P_{0n} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_n q_0}{\Sigma p_0 q_n}\right]} \times 100$$

$$= \frac{690}{580} \times \frac{1150}{960} \times 100 \qquad = \sqrt{[1.189 \times 1.197]} \times 100$$

$$= \mathbf{119.29.}$$

Example 10 : From the following data calculate price index numbers for the year 1980 with 1970 as the base year by using (i) Laspeyre's method (ii) Paashe's method (iii) Marshall Edgeworth method (iv) Fisher's method.

Commodity	**1970**		**1980**	
	Price(P_0)	**Quantity(q_0)**	**Price(P_n)**	**Quantity(q_n)**
A	20	8	40	6
B	50	10	60	5
C	40	15	50	15
D	20	20	20	25

Solution :

Construction of Price Index Numbers

Commodity	**1970**	**1980**				

	P_0	q_n	P_n	q_n	P_0q_n	P_nq_0	P_0q_n	P_nq_0
A	20	8	40	6	160	120	320	240
B	50	10	60	5	500	250	600	300
C	40	15	50	15	600	600	750	750
D	20	20	20	25	400	500	400	500
Total					$\sum p_0q_0$ =1660	$\sum p_0q_n$ =1170	$\sum p_nq_0$ =2070	$\sum p_nq_n$ =1790

(i) On using formula (1), Laspoyre's price index will be :

$$= P_{0n} = \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100 \qquad = \frac{2070}{1660} \times 100 = \mathbf{124.7}$$

(ii) On using formula (2), Pasche's price index becomes

$$= P_{0n} = \frac{\sum p_n q_n}{\sum p_0 q_n} \times 100 \qquad = \frac{1790}{1470} \times 100 = \mathbf{121.77}$$

(iii) Onusing formula (4), Marshall-Edgeworth price index can be computed as

$$= P_{0n} = \frac{\sum p_n(q_0+q_n)}{\sum p_0(q_0+q_n)} \times 100 \qquad = \frac{\sum p_n q_0 + \sum p_n q_0}{\sum p_0 q_0 + \sum p_0 q_n} \times 100$$

$$= \frac{2070+1790}{1660+1470} \times 100 \qquad = \frac{3860}{3130} \times 100 = \mathbf{123.32}$$

(iv) On using formula (7), Fisher's price index is

$$P_{0n} = \sqrt{\left[\frac{\sum p_n q_0}{\sum p_0 q_0} \times \frac{\sum p_n q_n}{\sum p_n 0 q_n}\right]} \times 100 = \sqrt{[1.247 \times 1.2177]} \times 100$$

$$= \mathbf{123.33.}$$

6.1.3 Weighted Average of Relative Method

As discussed earlier, weighted average of relative method is used to overcome the limitations of the simple average of relatives. Weighted arithmetic mean is the most common weighted average for computing index number, although weighted geometric mean can also be used. The general formula for weighted average of relatives can be written as –

$$P_{0n} = \frac{\sum\left(\frac{P_n}{P_0} \times 100\right) W}{\sum W} \times 100 = \frac{\sum RW}{\sum W} \quad .... (8)$$

$$\text{Where } R = \left(\frac{P_n}{P_0} \times 100\right) \quad .... (9)$$

Similarly, the index number based on geometric mean of price relatives is

$$P_{on} = \text{Antilog}\left(\frac{\sum W \log R}{\sum W}\right)$$

In this method, however each price relative is generally weighted by the total of the commodity in terms of some monetary value, say in Rs. etc. The value of the commodity is obtained by multiplying the price p of the commdity by the quantity q. Thus, the weight are given by the value, i.e., W = pq. Depending
on whether the base year values p_0q_0 are taken as weights, the price index formulae obtained by using weighted average of relatives are -

1. Weight arithmetic mean of price relatives using base year value weight is :

$$P_{0n} = \frac{\sum\left(\frac{P_n}{P_0}\right)(p0q0)}{\sum p0q0} \times 100 = \frac{\sum P_n q_0}{\sum P_0 q_0} \quad .... (10)$$

2. Weighted arithmetic mean of price relative using current year value weights is -

$$P_{0n} = \frac{\Sigma\left(\frac{P_n}{P_0}\right)(pnqn)}{\Sigma pnqn} \times 100 \qquad \text{.... (11)}$$

$$\text{or } P_{0n} = \frac{\Sigma\left(\frac{P_n}{P_0}\right) W}{\Sigma W} \times 100$$

Example 11 : Construct an index number for the following data using weighted average of price relative method.

Commudity	Base year Price (Rs.)(P.)	Current year Price(Rs.)(P.)	W eights
A	42.5	52.0	30
B	29.5	37.5	40
C	21.5	19.5	15
D	88.5	81.0	15

Solution : Computing weighted average of price relative index.

Commudity	Base year Price (Rs.)(P.)	Current year Price(Rs.)(P.)	Weights
A	42.5	52.0	30
B	29.5	37.5	40
C	21.5	19.5	15
D	88.5	81.0	15

Solution : Computing weighted average of price relative index.

Commudity	Base year Price (Rs.)P.	Current year Price(Rs.)P.	Weights $R=p_n/p_0$ X 100	Weights W	RW
A	42.5	52.0	122.4	30	3672.0
B	29.5	37.5	127.1	40	5084.0
C	21.5	19.5	90.7	15	1360.5
D	88.5	81.0	91.5	15	1372.5
				ΣW=100	ΣRW-11489.0

Thus, the weighted average of price relative index

$$P_{0n} = \frac{\Sigma\left(\frac{P_n}{P_0}\right) W}{\Sigma W} \times 100 = \frac{\Sigma RW}{\Sigma W} = \frac{11489.0}{100}$$

$$= 114.89$$

Example 12 : An enquiry into the family budget of middle class family gave the following information -

Item	Food	Rent	Clothing	Fuel	Others
Expenditure % :	30%	15%	20%	10%	25%
Price (Rs.) in 1985 :	100	20	70	20	40
Price (Rs.) in 1986	90	20	60	15	35

Compute the price index for 1986 by using (1) weighted arithmetic mean of price relatives (ii) weighted geometric mean of price relatives.

Solution : **Computation of weighted index**

Item	Weight W	P0	Pn	Price Relative $R=P_n/P_0X100$	WR	Log R	W log R
Food Rent	30 15	100 20	90 20	90.0 100.0	2700.00 1500.00	1.9542 2.0000	58.626 30.000
Clothing	20	7 0	60	85.7	1714.00	1.9330	38.660
Fuel	10	2	15	75.0	750.0	1.8751	18.751

		0					
Others	25	40	55	137.5	3437.50	2.1383	53.457
	$\sum W=$ 100				$\sum WR=$ 10101.50		$\sum W \log R=$ 199.494

(i) Index based on weighted A.M. of relatives :

$$P_{0n} = \frac{\sum RW}{\sum W} = \frac{10101.5}{100} = \mathbf{102.02}$$

(i) Index based on weighted G.M. of relatives :

$$P_{0n} = \text{antilog}\left(\frac{\sum W \log R}{\sum W}\right) = \text{antilog}\left(\frac{199.494}{100}\right) = \text{antilog}\,[1.99494]$$

$$= \mathbf{98.33}$$

Example 13 : From the information on the next page construct index number of 1990 by the method of weighted average of relavites using (i) base year value as weights (ii) given year values weights.

Article	1985		1990	
	Price(P_0)	Quantity(q_0)	Price(P_n)	Quantity(q_n)
A	8	50	20	40
B	6	10	18	2
C	4	5	5	2

Solution: Construction of Price Index (weighted average relative method).

Article	1985		1990						
	P_0	q_0	P_n	q_n	$V_0=P_0q_0$	$V_n=P_nq_n$	$R=P_n/P_0$	RV_0	RV_n
A	8	50	20	40	400	800	2.5	1000.0	2000.0
B	6	10	18	2	60	36	3.0	180.0	108.0
C	4	5	5	2	20	16	2.0	40.0	32.0
Total					$\sum P_0q_0 = 480$	$\sum P_nq_n = 852$		$\sum RV_0 = 2070$	$\sum RV_n = 1790$

(i) Price index when base year value weights are used:

$$P_{0n} = \frac{\sum\left(\frac{P_n}{P_0}\right)(p0q0)}{\sum p0q0} \times 100 = \frac{\sum RV_0}{\sum V_0} \times 100 = \frac{1220.0}{480.0} \times 100$$

= 473.45.

(ii) Price index when given year value weights are used:

$$P_{0n} = \frac{\sum\left(\frac{P_n}{P_0}\right)(pnqn)}{\sum pnqn} \times 100 = \frac{\sum RV_n}{\sum V_n} \times 100 = \frac{\sum RV_n}{\sum V_n} \times 100$$

= 473.45.

Quantity or Volume Index numbers

Instead of comparing price changes over a period of time, we may be interested in analysing changes in quantity of production or consumption of certain commodities over a given period of time. In order to study such changes, we construct quantity index numbers. In this regard, a few quantity index numbers (Q_{0n}) formula may be listed as under.

(i) Simple aggregative quantity index :

$$Q_{0n} = \frac{\sum q_n}{\sum q_0} \times 100 \qquad \text{... (12)}$$

(ii) Simple average of relatives method :

$$Q_{0n} = \frac{\sum (q_n/q0)}{N} \times 100 \qquad \text{... (13)}$$

(iii) Laspeyres' Quantity index : Using base year prices as weights, Laspeyres' quantity index is given by

$$Q_{0n} = \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100 \qquad \text{... (14)}$$

(iv) Paasche's quantity index : Using given year prices as weights, Paasche's quantity index can be formulate as

$$Q_{0n} = \frac{\sum p_n q_n}{\sum p_0 q_n} \times 100 \qquad \text{... (15)}$$

Flasher's and Marshall-Edgeworth's formula for quantity index numbers can also be obtained on the similar pattern.

While the price index number measure the change in the value of a fixed aggregate of goods at varying prices, the quantity index number measures the change in value of a varying aggregate of goods at fixed prices. Thus, the price index number enables us to know the amount of expenditure in a given year if the same volume of quantity is consumed at varying prices. On the other hand, the quantity index number tells us how much we shall spend in the given year if varying quantities of commodities are bought at the same price.

Example 14 : Compute (i) Laspeyres' (ii) Paasche's and (iii) Fisher's quantity index numbers from the following data :

Article	1982		1984	
	Price(P_0)	Quantity(q_0)	Price(P_n)	Quantity(q_n)
A	5	10	4	12
B	8	6	7	7
C	6	3	5	4

Solution : Computation of Price Index Numbers

Commodity	1982		1984					
	P_0	q_n	P_n	q_n	P_0q_n	P_nq_0	P_0q_n	P_nq_0
A	5	10	4	12	50	60	40	48
B	8	6	7	7	48	56	42	49
C	6	3	5	4	18	24	15	20
Total					$\sum p_0q_n$ = 116	$\sum p_nq_0$ = 140	$\sum p_0q_n$ = 97	$\sum p_nq_0$ =117

Using (14), Laspeyres' quantity index is -

$$= Q_{0n} = \frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times 100 = \frac{97}{116} \times 100 = 83.62$$

Using (23), Paasche's quantity index is -

$$= Q_{0n} = \frac{\Sigma p_n q_n}{\Sigma p_0 q_n} \times 100 = \frac{117}{140} \times 100 = 83.57$$

Finally, Fiasher's Quantity Index will be

$$Q_{0n} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_n q_n}{\Sigma p_n q_n}\right]} \times 100$$

$$= \sqrt{\left[\frac{97}{116} \times \frac{117}{140}\right]} \times 100 = \sqrt{\left[\frac{11349}{13697}\right]} \times 100 = 91.02$$

Example 15 : Use the following data to find Fisher's price and quantity indices-

Article	Base year		Current year	
	Price	Quantity	Price	Quantity
A	6	50	10	56
B	2	100	2	120
C	4	60	6	60
D	10	30	12	24
E	8	40	12	36

Solution : Computation of Fisher's price and quantities indices.

Article	Base year		Current year					
	P_0	q_n	P_n	q_n	P_0q_n	P_nq_0	P_0q_n	P_nq_0
A	6	50	10	56	300	336	500	560
B	2	100	2	120	200	240	200	240
C	4	60	6	60	240	240	360	360
D	10	30	12	24	300	240	360	288
E	8	40	12	36	320	288	480	432
Total					$\sum p_0q_0$ =1360	$\sum p_0q_n$ =1344	$\sum p_nq_0$ =1900	$\sum p_nq_n$ =1880

Fisher's Price Index $P_{0n} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_n q_n}{\Sigma p_0 q_n}\right]} \times 100$

$$= \sqrt{\left[\frac{1900}{1360} \times \frac{1880}{1344}\right]} \times 100$$

$=\sqrt{[1.95219]} \times 100 =$ **139.79**

Fisher's Quantity Index $Q_{0n} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_n q_n}{\Sigma p_n q_n}\right]} \times 100$

$$= \sqrt{\left[\frac{1900}{1360} \times \frac{1880}{1344}\right]} \times 100$$

$=\sqrt{[0.9778328]} \times 100 = 0.9889 \times 100 =$ **98.89**

6.2 Tests of Adequacy of Index Numbers

Earlier, while discussing the relatives, we observed that relatives follow certain properties. When these properties are true for an individual commodity, these should also hold good for a group of commodities. As such, the index number as an aggregative relative should also satisfy these properties. In view of this, we now discuss these properties in the light of various index number formulae. A good index number should satisfy the following tests or properties.

1. Unit Test
2. Time Reversal Test
3. Factor Reversal Test
4. Circular Test

Unit Test

According to unit-test, an index number should be independent of the unit in which prices and quantities of various commodities and quoted. This test is satisfied by all the index number formulae except the simple aggregative index.

Time Reversal Test

According to Prof. Fisher "the formula for calculating an index number should be such that it gives the same ratio between one point of comparison and the other, no matter which of the two is taken as base." In other words, if the two periods, the base and the reference period are interchanged, the product of the two index numbers should be unity, i.e, they should be reciprocal of each other. Symbolically,

$$P_{0n} \times P_{n0} = 1 \text{ or } \quad P_{0n} = 1/P_{n0} \qquad \ldots (16)$$

An index formula satisfying the criteria in equation (16) is said to satisfy the time reversal test. Let us see whether Laspeyres' index number satisfy this property or not. For which, Laspeyres' formula is -

$$= P_{0n} = \frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times 100$$

$$= P_{n0} = \frac{\Sigma p_n q_n}{\Sigma p_0 q_n} \times 100$$

$$\text{Now} \quad = P_{0n} X P_{no} = \frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times 100 \times \frac{\Sigma p_n q_n}{\Sigma p_0 q_n} \times 100 = 1.$$

So Laspeyres' formula does not satisfy the time reversal test. Similarly, Paasche's index formula does not satisfy this test. However, if we consider Fisher's index

$$P_{0n} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_n q_n}{\Sigma p_n q_n} \right]}$$

$$\text{and thus} \quad P_{0n} = \sqrt{\left[\frac{\Sigma p_o q_n}{\Sigma p_n q_n} \times \frac{\Sigma p_0 q_0}{\Sigma p_n q_0} \right]}$$

$$\text{Now } P_{on} \text{ X } P_{0n} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \text{ X } \frac{\Sigma p_n q_n}{\Sigma p_n q_n}\right]} \text{ X } \sqrt{\left[\frac{\Sigma p_0 q_n}{\Sigma p_n q_n} \text{ X } \frac{\Sigma p_0 q_0}{\Sigma p_n q_0}\right]}$$

Fisher's Index satisfies time reversal test.

Similarly, students can see that **Marshall-Edgeworth's** and **Walsh's** index numbers also satisfy time reversal tests.

Factor Reversal Test

We know that if the two factors p and q is inter-changed in a price index formula (P_{0n}), we get quantity index formula (Q_{0n}). Then we expect that the product of P_{0n} and Q_{0n} should be equal to the true value ratio. Index should give the true ratio of value in the given year (n) to the value in the base year (0). Symbolically, the factor reversal test is satisfied if

$$P_{0n} \text{ X } Q_{0n} = \frac{\Sigma p_n q_n}{\Sigma p_0 q_n} = \frac{\text{Value in the given year (n)}}{\text{Value in the base year (o)}} \qquad ...(17)$$

Now let us consider the case of Laspeyres' Index for which

$$P_{0n} = \frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \quad \text{and} \quad Q_{0n} = \frac{\Sigma q_n p_0}{\Sigma q_0 p_0}$$

$$\text{Thus,} = P_{0n} \text{ X } Q_{0n} = \frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \text{ x } \frac{\Sigma q_n p_0}{\Sigma q_0 p_0} = \frac{\Sigma p_n q_n}{\Sigma p_0 q_0}$$

Thus, **Laspeyre's index** does not satisfy factor reversal test. However, in the case of Fisher's index, we have,

$$P_{on} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \text{ X } \frac{\Sigma p_n q_n}{\Sigma p_n q_n}\right]} \text{ and } Q_{0n} \sqrt{\left[\frac{\Sigma q_n p_n}{\Sigma q_0 p_0} \text{ X } \frac{\Sigma q_n p_n}{\Sigma q_n p_n}\right]}$$

$$P_{on} \times Q_{0n} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_n q_n}{\Sigma p_n q_n}\right]} \times \sqrt{\left[\frac{\Sigma q_0 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_n p_n}{\Sigma q_n q_n}\right]}$$

$$= \frac{\Sigma p_n q_n}{\Sigma p_0 q_n}$$

Thus, **Fisher's index also satisfies the factor reversal test.** Now since Fisher's index number is one which satisfies both the time reversal and factor reversal test, so it is called **Fisher's ideal index number.**

Circular Test:

This is another test for the adequacy of the index number. This is based on the shifting of the base period and thus is an extension on the time reversal test. According to this test, an index number should also work in a circular way, i.e. the test requires that

$$P_{01} \times P_{12} \times P_{20} = 1 \qquad \ldots(18)$$

For, four time points, we have $P_{01} \times P_{12} \times P_{23} \times P_{30} = 1$

In particular, for two time points, 0 and n, one gets

$$P_{0n} \times P_{n0} = 1$$

which is nothing but time reversal test discussed above. It can be observed that none .... weighted index numbers satisfies this test. However, for three 0, 1 and 2, if the test is applied to index obtained by simple aggregate method, one gets -

$$P_{01} X\ P_{12} X\ P_{20} = \frac{\Sigma p_1}{\Sigma p_0} \times \frac{\Sigma p_2}{\Sigma p_1} \times \frac{\Sigma p_0}{\Sigma p_2} = 1$$

Similarly, if we apply circular test to index numbers obtained by fixed weight aggregative method, we get -

$$P_{01} X\ P_{12} X\ P_{20} = \frac{\Sigma p_1 q}{\Sigma p_0 q} \times \frac{\Sigma p_2 q}{\Sigma p_1 q} \times \frac{\Sigma p_0 q}{\Sigma p_2 q} = 1.$$

Thus, for index obtained by simple aggregate method or by fixed weight aggregative method, the circular test is satisfied.

Example 16. Calculate Laspeyre's and Paasche's price indices for the year 1980 from the following data. Prove that both the formulae do not satisfy the Time Reversal Test.

Commodity	Price (Rs.)		Quantities (kgs)	
	1979	1980	1979	1980
A	2.0	2.50	3	5
B	2.5	3.00	4	6
C	3.0	2.50	2	3
D	1.0	0.75	1	2

Solution : Computation of Laspeyre's and Paasche's indices

Article	Base year		Current year					
	P_0	q_n	P_n	q_n	P_0q_n	P_nq_0	P_0q_n	P_nq_n
A	2.0	3	2.50	5	6.0	7.5	10.0	12.5
B	2.5	4	3.00	6	10.0	12.00	15.0	18.0
C	3.0	2	2.50	3	6.0	5.00	9.0	7.5
D	1.0	1	0.75	2	1.0	0.75	2.0	1.5
Total					$\sum p_0q_0$ =23.0	$\sum p_nq_n$ =25.25	$\sum p_0q_n$ =36.0	$\sum p_nq_n$ =39.5

$$\textbf{Laspeyrs' index} = P_{0n} = \frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times 100 = \frac{25.25}{23.00} \times 100 = 109.78$$

$$\textbf{Paasche's index} = P_{0n} = \frac{\Sigma p_n q_n}{\Sigma p_0 q_n} \times 100 = \frac{39.50}{36.00} \times 100 = 109.72$$

Time Reversal Test : This test is satisfied if $P_{0n} \times P_{n0} = 1$

$$P_{0n} = \frac{\Sigma p_n q_0}{\Sigma p_0 q_n} \text{ and } P_{n0} = \frac{\Sigma p_n q_0}{\Sigma p_n q_0}$$

$$\therefore \quad P_{0n} \times P_{n0} = \frac{\Sigma p_n q_n}{\Sigma p_0 q_n} \times \frac{\Sigma p_0 q_0}{\Sigma p_n q_0} = \frac{39.50}{36.00} \times \frac{23.00}{25.25} = 1$$

Thus, Paasche's indedx also does not satisfy the Time Reversal Test.

Example 17. The following table gives the prices and quantities of 5 commodities in the base and current year. Use it to verify whether Fisher's ideal index satisfies the time reversal test.

Commodity	Base year		Current year	
	Unit price (Rs.)	Quantity (kgs.)	Unit price (Rs.)	Quantity (kgs.)
A	5	50	5	70
B	5	75	10	80
C	10	80	12	100
D	5	20	8	30
E	10	50	5	60

Solution : Computation Fisher's Ideal Index

Commodity	Base year		Current year					
	P_0	q_0	P_n	q_n	$P_0 q_0$	$P_n q_n$	$P_n q_n$	$P_n q_n$
A	5	50	5	70	250	350	250	350
B	5	75	10	80	375	400	750	800
C	10	80	12	100	800	1000	960	1200
D	5	20	8	30	100	150	160	240
E	10	50	5	60	500	600	250	300
Total					$\sum p_0 q_0$ =2025	$\sum p_n q_n$ =2500	$\sum p_0 q_n$ =2370	$\sum p_n q_n$ =2890

Fisher's ideal index $= P_{0n} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_n q_n}{\Sigma p_0 q_n}\right]}$

$$= \sqrt{\left[\frac{2370}{2025} \times \frac{2890}{2500}\right]} \times 100$$

$=\sqrt{[1.352948]} \times 100 = 1.1632 \times 100 =$ **116.32**

For time reversal test -

$$P_{on} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_n q_n}{\Sigma p_0 q_n}\right]} = \sqrt{\left[\frac{2370}{2025} \times \frac{2890}{2500}\right]}$$

$$P_{on} = \sqrt{\left[\frac{\Sigma p_0 q_n}{\Sigma p_n q_n} \times \frac{\Sigma p_0 q_0}{\Sigma p_n q_0}\right]} = \sqrt{\left[\frac{2500}{2890} \times \frac{2025}{2370}\right]}$$

$$P_{on} \times P_{no} = \sqrt{\left[\frac{2370}{2025} \times \frac{2890}{2500}\right]} = \sqrt{\left[\frac{2500}{2890} \times \frac{2025}{2370}\right]} = 1.0$$

Thus, Fisher's Index satisfies the time reversal test.

Example 18. In the above example, verify that Fisher's index also satisfies the factor reversal test.

Solution : The factor reversal test is satisfied if

$$P_{0n} \times Q_{0n} = \frac{\Sigma p_n q_n}{\Sigma p_0 q_o}$$

Let us consider -

$$P_{on} = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_n q_n}{\Sigma p_0 q_n}\right]} = \sqrt{\left[\frac{2370}{2025} \times \frac{2890}{2500}\right]}$$

$$Q_{on} = \sqrt{\left[\frac{\Sigma q_n p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_n p_n}{\Sigma q_0 p_n}\right]} = \sqrt{\left[\frac{2500}{2025} \times \frac{2890}{2370}\right]}$$

$$\therefore P_{on} \times Q_{on} = \sqrt{\left[\frac{2370}{2025} \times \frac{2890}{2500}\right]} = \sqrt{\left[\frac{2500}{2025} \times \frac{2890}{2370}\right]}$$

$$= \frac{2890}{2025} = \frac{\Sigma p_n q_n}{\Sigma p_0 q_o}$$

Thus, Fisher's index also satisfies the factor reversal test.

Example 19. From the following prove that Fisher's ideal index satisfies both the time reversal and the factor reversal test -

Commodity	Base year		Current year	
	Price	Quantity	Price	Quantity
A	6	50	10	60
B	2	100	2	120
C	4	60	6	60

Solution : Computation of Fisher's Ideal Index Number

Commodity	Base year		Current year					
	P_0	q_0	P_n	q_n	P_0q_0	P_0q_n	P_nq_0	P_nq_n
A	6	50	10	60	300	360	500	600
B	2	100	2	120	200	240	200	240
C	4	60	6	60	240	240	360	360
Total					$\Sigma p_0q_0 = 700$	$\Sigma p_0q_n = 840$	$\Sigma p_nq_0 = 1000$	$\Sigma p_nq_n = 1200$

$$\text{Fisher's index (Price)} = P_{0n} = \sqrt{\left[\frac{\Sigma p_nq_0}{\Sigma p_0q_0} \times \frac{\Sigma p_nq_n}{\Sigma p_0q_n}\right]} \times 100$$

$$= \sqrt{\left[\frac{1000}{700} \times \frac{1200}{840}\right]} \times 100 = \mathbf{142.86}$$

Time Reversal Test. This test is satisfied if $P_{0n} \times P_{n0} = 1$.

$$\text{Here } P_{on} = \sqrt{\left[\frac{\Sigma p_nq_0}{\Sigma p_0q_0} \times \frac{\Sigma p_nq_n}{\Sigma p_0q_n}\right]} = \sqrt{\left[\frac{1000}{700} \times \frac{1200}{840}\right]}$$

$$\text{and } P_{no} = \sqrt{\left[\frac{\Sigma p_0q_n}{\Sigma p_nq_n} \times \frac{\Sigma p_0q_0}{\Sigma p_nq_0}\right]} = \sqrt{\left[\frac{700}{1000} \times \frac{840}{1200}\right]}$$

$$\therefore P_{on} \times P_{no} = \sqrt{\left[\frac{1000}{700} \times \frac{1200}{840}\right]} \times \sqrt{\left[\frac{700}{1000} \times \frac{840}{1200}\right]} = 1.00$$

Thus, Fisher's Index satisfies the **Time Reversal Test.**

Factor Reversal Test : The test is satisfied if - $P_{on} \times Q_{on} = \frac{\Sigma P_nQ_n}{\Sigma P_oQ_o}$

$$\text{Here } P_{on} = \sqrt{\left[\frac{\Sigma p_nq_0}{\Sigma p_0q_0} \times \frac{\Sigma p_nq_n}{\Sigma p_0q_n}\right]} = \sqrt{\left[\frac{1000}{700} \times \frac{1200}{840}\right]}$$

$$\text{and } Q_{0n} = \sqrt{\left[\frac{\Sigma q_np_0}{\Sigma q_0p_0} \times \frac{\Sigma q_np_n}{\Sigma q_0p_n}\right]} = \sqrt{\left[\frac{840}{700} \times \frac{1200}{1000}\right]}$$

$$\therefore P_{on} \times Q_{on} = \sqrt{\left[\frac{1000}{700} \times \frac{1200}{840}\right]} \times \sqrt{\left[\frac{840}{700} \times \frac{1200}{1000}\right]}$$

$$= \frac{1200}{700} = \frac{\Sigma p_nq_n}{\Sigma p_0q_o}$$

Thus, Fisher's index also satisfies **Factor Reversal Test.**

Link and Chain Relatives

Let $p_1, p_2, p_3,$.... denote prices of a commodity during successive intervals of time 1, 2, 3, ...respectively then the price relatives of each time interval with respect to the preceding time interval as base, i.e., p_{12}, p_{23}, p_{34},. , etc., are called **link relatives.** Now using this series of price relatives and circular property of relatives, one can get the price relative for a given period with respect to any other period as base. To put it other way, let us consider the modified circular property of price relatives which states

$p_{13} = p_{12}, p_{23}$ .... (19)

or $p_{14} = p_{12}, p_{23}, p_{34}$ .... (20)

That is, on using (19), the price relative of period 3 with respect to period 1 as base can be obtained from the link relative p_{12} and p_{23}. Similarly, link relatives p_{12},p_{23}, and p_{34} can be used to get p_{14}, i.e., the price relative of period 4 with respect to period 1as base.

Thus, the price relatives with respect to a fixed base period can be obtained by using link relatives. These link relatives of one period with respect to preceding as base are chained together by successive multiplication to get a chain price relative of a period with respect to a fixed base period. In view of this, it is called a chain relative or chain index. The concept of link and chain relatives can be similarly used in the case of quantity and value relatives.

Illustration 20. If the prices of a commodity during 1984, 1985, 1986 and

1987 are Rs. 30, 35, 38 and 42 respectively, then the price link relatives can be obtained as shown in the table -

Year	1984(1)	1985(2)	1986(3)	1987(4)
Price(Rs.)	30(p_1)	35(p_2)	38(p_3)	42(p_4)
Link relatives		$P_{12} = \frac{P_2}{P_1} = \frac{35}{30}$	$P_{23} = \frac{P_3}{P_2} = \frac{38}{35}$	$P_{34} = \frac{P_4}{P_3} = \frac{42}{38}$

Now on using these link relative and their circular property, the price relative with respect to a fixed base can be obtained. For example, the price relative for the year 1986 with base year 1984 can be enumerated as

$$P_{13} = P_{12} \text{ X } P_{23} = \frac{35}{30} \quad \frac{38}{35} = \frac{38}{30} = \mathbf{126.67\%}$$

Similarly, the price relative of 1987 with base year as 1984 becomes.

$$P_{14} = P_{12} \text{ X } P_{23} \text{ X } P_{34} = \frac{35}{30} \text{ X } \frac{38}{35} \text{ X } \frac{42}{38} = \frac{42}{30} = \mathbf{140\%}$$

Thus, the construction of chain indices involves the following steps-

1. Express the figures for each period as percentage of the preceding period to get the link relatives (L.R.)
2. These link relatives are chained together by successive multiplication to get the

chain indices or relatives of a period with respect to a fixed base period Symbolically, the chain relatives or indicies are computed as

P_{01} = First link relative P_{02}

$= P_{01} X P_{12}$

$P_{03} = P_{01} X P_{12} X P_{23} = P_{02} X P_{23}$

$P_{01} = P_{01} X P_{12} X P_{23} X P_{34} = P_{23}$

$P_{0k} = P_{01} X P_{12} X ... X P_{(k-1)} = P_{0(k-1)} X P_{(k-1)k}$

or use the following formula formula for computing chain index (C.I.)

$$C.I. = \frac{\text{Current year L.R. X Preceeding C.I.}}{100} \quad ...(21)$$

It is notable here that the meanings of the terms fixed based index (F.B.I), link relative or index (L.R.). chain index (C.I.)and chain base index (C.B.I.), should be clearly understood. In these terms link relative (L.R.) and chain base index (C.B.I.) convey the same meaning. While, for an index series, chain indies are the same as the fixed base index.

Further, in some specific situations, we also need to convert chain base index numbers (C.B.I.) to fixed base index (F.B.I.) for which the following formula is applicable -

$$\text{Current year F.B.I.} = \frac{\text{Current year C.B.I. X Previous year F.B.I.}}{100} \quad ...(22)$$

The computation procedure will be more clear from the following examples-

Example 20. From the following data of wholesale prices of a certain commodity, construct (i) Fixed Base Index (1979 = 100) and (ii) Chain index numbers.

Year :	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
Price :	75	50	65	60	72	70	69	75	84	80

Solution : Using the formula-

(i) **Fixed Base Index of a given year** $= \frac{\text{Price of the given year}}{\text{Price of the base year}} 100$

$$\text{(ii) Chain Index} = \frac{\text{Link relative of the given year X Chain index of the preceeding year}}{100}$$

Using formula (i) and (ii), the fixed base and chain indices are given in the following table -

Year	Price	Fixed Base Index (1979=base)	Link Relatives (L.R.)	Chain Index (C.L.)
1979	75	100	100	100
1980	50	$\frac{50}{75}$ X 100=66.67	$\frac{50}{75}$ X 100=66.67	$\frac{66.67 \text{ X } 100}{100}$ =66.67
1981	65	$\frac{65}{75}$ X 100=86.67	$\frac{65}{50}$ X 100=130.00	$\frac{130 \text{ X } 66.67}{100}$ =86.67
1982	60	$\frac{60}{75}$ X 100=80.00	$\frac{60}{65}$ X 100=66.67	$\frac{66.67 \text{ X } 86.67}{100}$ =80.00
1983	72	$\frac{72}{75}$ X 100=96.00	$\frac{72}{60}$ X 100=120	$\frac{120 \text{ X } 80}{100}$ =96.00
1984	70	$\frac{70}{75}$ X 100=93.33	$\frac{70}{72}$ X 100=97.22	$\frac{97.22 \text{ X } 96}{100}$ =93.33
1985	69	$\frac{69}{75}$ X 100=92.00	$\frac{69}{70}$ X 100=98.57	$\frac{98.57 \text{ X } 93.33}{100}$ =92.00
1986	75	$\frac{75}{75}$ X 100=100.00	$\frac{75}{69}$ X 100=108.69	$\frac{108.69 \text{ X } 92}{100}$ =100
1987	84	$\frac{84}{75}$ X 100=112.00	$\frac{84}{75}$ X 100=112.00	$\frac{112 \text{ X } 100}{100}$ =112
1988	80	$\frac{80}{75}$ X 100=106.67	$\frac{80}{84}$ X 100=95.24	$\frac{95.24 \text{ X } 112}{100}$ =106.67

Remark : It may be noted that the chain indices are the same as the fixed base indx numbers.

Example 21. Use the following chain base index numbers (C.B.I.) to obtain the fixed base indexnumbers.

Year	:	1978	1979	1980	1981	1982	1983
Chain Base Index	:	105	75	71	105	95	90

Solution : Using formula (22), the required incides are shown in table.

Computation of fixed bas index numers

Year	Chain base Index (C.B.I.)	Fixed base Index number (F.B.I.)
1978	105	= 105.5
1979	75	$\frac{75 \text{ X } 105}{100}$ X 78.75
1980	71	$\frac{71 \text{ X } 78.75}{100}$ X 55.91
1981	105	$\frac{105 \text{ X } 55.91}{100}$ X 58.71
1982	95	$\frac{95 \text{ X } 58.71}{100}$ X 55.77
1983	90	$\frac{90 \text{ X } 55.77}{100}$ X 50.20

Example 22. Compute chain Index numbers with 1986 prices as base from the following table giving the average wholesale price of the commodities A,B,C for years 1987 to 1990.

Commodities	1986	1987	1988	1989	1990
A	20	16	28	35	21
B	25	30	24	36	45
C	20	25	30	24	30

Solution : Computation of Chain indicies

Commodity	Price relative based on precending year				
	1986	**1987**	**1988**	**1989**	**1990**
A	100	$\frac{16}{20}$ x 100=	$\frac{28}{16}$ x 100=	$\frac{35}{28}$ x 100	$\frac{21}{35}$ x 100=
B	100	$\frac{30}{25}$ x 100=	$\frac{24}{30}$ x 100=	$\frac{36}{24}$ x 100	$\frac{45}{36}$ x 100=
C	100	$\frac{25}{20}$ x 100=	$\frac{30}{25}$ x 100=	$\frac{24}{30}$ x 100	$\frac{30}{24}$ x 100=

Total of link Relatives	300	325	375	355	310
Average of link relatives	100	108.33	125	118.33	103.33
Chain Relatives	100	$\frac{108.33 \times 100}{100}$ = 108.33	$\frac{125 \times 108.33}{100}$ = 135.41	$\frac{118.33 \times 135.41}{100}$ = 160.23	$\frac{103.33 \times 160.23}{100}$ = 165.5

Example 22. From the chain base index given below, perpare fixed base index numbers.

Year :	1981	1982	1983	1984	1985
Index :	110	160	140	200	150

Solution : Conversion from chain base to fixed base index

Year	Chain base index	Conversion	Fixed index
1981	110		110.0
1982	160	(160x110)/100	176.0
1983	140	(140x176)/100	246.4

1984	200	(200x264.4)/100	492.8
1985	150	(150x492.8)/100	739.2

Example 24. From the following fixed base number index (F.B.I.) prepare chain base indes (C.B.I.) :

Year :	1980	1981	1982	1983	1984	1985
(C.B.I.) :	220	250	300	280	350	415

Solution :

Conversion from F.B.I. to C.B.I.

Year	Chain base index	Conversion	Fixed index
1980	220	-	100.00
1981	250	(250x220)/100	113.64
1982	300	(300x250)/100	120.00
1983	280	(280x300)/100	93.33
1984	350	(350x280)/100	125.00
1985	415	(415x350)/100	118.57

Base Shifting, Splicing and Deflating of Index Numbers Base Shifting:

In reference to index numbers, base shifting means to prepare a new index series with a new base period in place of an old one. Thus, a base shifting is nothing but a procedure of recasting an index series by shifting its base period to some recent or more relevant base period. Base shifting is necessary in the following situations -

1. When the base period of the index series is too old or too distant from the current period.

2. When we wish to compare two or more index series with different base periods then, for making valid comparisons, it is necessary that the given index series be expressed with a common base period.

In base shifting procedure, the index number of the new base year is taken as 100 and then the remaining index numbers in the series are expressed as percentage of the index number selected as new base.

Thus, the index series may be recast by using the following formula -

$$\text{Recasted Index No. of any year} = \frac{\text{old Index No. of the year}}{\text{Index No. of new base year}} \times 100 \quad \text{..............(23)}$$

Example 25. Assuming 1979 as the base prepare new index numbers from the indices given below :

Year	:	1976	1977	1978	1979	1980
Indices	:	100	110	125	250	300

Base shifting from 1976 to 1978

Year	Indices (Bease 1976)	Base Shifting	New Indices (base 1979)
1976	110	(100 X 100)/250	40
1977	110	(110x100)/250	44
1978	175	(175x100)/250	70
1979	250	(250x100)/250	100
1980	250	(300x100)/250	120

Example 26. Reconstruct the following index series using 1980 as base :

Year	:	1976	1977	1978	1979	1980	1981	1982
Index No.	:	110	130	150	175	180	200	220

Solution :

Base Shifting

Year	Indices No.	Base Shifting	New Indices (base 1980)
1976	110	(110 X 100)/180	61.11
1977	130	(130x100)/180	72.22
1978	150	(150x100)/180	83.33
1979	175	(175x100)/180	97.22

1980	180	(180x100)/180	100.00
1981	200	(200x100)180	111.11
1982	220	(220x100)180	122.22

Splicing

Splicing means combining two or more index series. For retaining continuity in comparison between two or more index series, we splice or combine them into a new single index series. For clarity, suppose there is an index series' A' with base period 1970 which discontinued in 1975 and then a new index series. 'B' is prepared with base period 1970 which was discontinued in 1975 and then a new index series 'B' is prepared with base period 1975. Then for comparing the two index series, we can splice them into a new continuous in index series in the following manner:-

1. Splice the index series B' to 'A' to obtain a new continuous index series with base 1970. This splicing procedure is known as forward splicing. In fact, in this splicing the base 1975 of index series 'B' has to be shifted to base 1970.

2. Splice the index series 'A' to 'B' to get a new continuous index series with base 1975, this splicing procedure is known as backward splicing. In other words, in this type of splicing the base period 1970 of index series 'A' is to be shifted to base 1975. Thus, the procedure in splicing is very much alike in that involved in base shifting.

The formula used for forward splicing is:-

$$\text{Required Index} = \frac{\text{Old index number on the existing base x Index number to be spliced}}{100} \quad \text{....(24)}$$

Also the formula used for backward splicing is :-

$$\text{Needed Index} = \frac{100}{\text{Old index number on the existing base x Index number to be spliced}} \quad \text{...(25)}$$

The following examples will clarify the procedure.

Example 27. Splice the following two index series, series A forward and the series B backwards.

Year	:	1983	1984	1985	1986	1987	1988
Series A	:	110	130	150			
Series B	:			100	110	140	150

Solution : **Splicing Two Index Series**

Year	Series		Index Numbers spliced forward to Series A	Index number splied backward to series B
	A	B		
1983	100			(100x100)/150=66.67
1984	130			(100x130)/150=86.67
1985	150	100	(150x100)/100=150	(100x150)/150=100
1986		110	(150x110)/100=165	
1987		140	(150x140)/100=210	
1988		150	(150x150)/100=225	

Thus,

Year	:	1983	1984	1985	1986	1980	1988
Forward Splice Series	:	100	130	150	165	210	225
Backward Splice Series	:	66.67	86.67	100	110	140	150

Example 28. We have the following three index series.

I-Series

Year	**:**	1980(Base)	1981	1982
Index No.	**:**	100	120	200

II-Series

Year	**:**	1982(Base)	1983	1984
Index No.	**:**	100	110	130

III-Series

Year	:	1984	1985	1986
Index No.	:	100	130	140

Obtain a continuous series with the base 1984 by splicing the three series.

Solution : **Splicing Three Series**

Year	Indices			Special Index Series (base 1984)
	I	II	III	
1980	100			(100x76.92)/200=38.10
1981	120			(120x76.92)/200=46.15
1982	200	100		(100x100)/130=76.92
1983		110		(110x110)/130=84.62
1984		130	100	(100x130)/130=100
1986			130	130
1986			140	140

Thus,

Year	:	1980	1981	1982	1983	1984	1985	1986
Continous Series	:	38.10	46.15	76.92	84.62	100	130	140

base(1984) Deflating :

Deflating means "making allowance for the effect of changing price level." An increase in price level of consumer goods over a period means are reduction in the purchasing power of the people. For example if the price of rice rises from Rs. 500 per quintal in 1990 to Rs. 1000 per quintal in 1992, this simply means that in 1992 the person can buy only half the amount of rice in 1992 if he decided to spent the same amount which he was spending in 1990. In other words, the value of the rupees is 50 paisa in 1992 as compared to that in 1990 i.e. the purchasing power of the money in 1992 is half as compared to that in 1990. Therefore, the purchasing power is given by the reciprocal of the index number and consequently the real income (or wages) is obtained by dividing the nominal income of the period by the corresponding index number and expressing the ratio in percentage. Thus

$$\text{Real wages} = \frac{\text{Money or normal wages}}{\text{Price Index}} \times 100$$

The real wages is also known as deflated wages (or income). In this way, we observe that, in the deflating procedure, the relevant price index is the deflator and a deflated value, in general, can be computed by useing the formula -

$$\textbf{Deflated value} = \frac{\text{Current Value}}{\text{Deflator}} \times 100$$

$$= \frac{\text{Current Value}}{\text{relevant index number}} \times 100$$

Example 29: The following table gives the money wages and cost of living index number based on 1979.

Year	:	1979	1980	1981	1982	1983	1984	1985
Forward Splice Series	:	65	70	75	80	90	100	120
Backward Splice Series	:	100	110	120	130	150	160	200

Solution :

$$\text{Real Wages} = \frac{\text{Normal Wages}}{\text{Index No.}} \times 100$$

Computation of Real Wages

Year	**Wages (Rs.)**	**Index (1979=100)**	**Deflected wages or Real Wages [Col(2)/Col.(3)]/100**
(1)	**(2)**	**(3)**	**(4)**
1979	65	100	(65x100)/100 = 65
1980	70	110	(70x100)/110 = 63.64
1981	75	120	(75x100)/120 = 62.50
1982	80	130	(80x100)/130 = 61.54
1983	90	150	(90x100)/150 = 60.00
1984	100	160	(120x100)/160 = 62.50
1985	120	200	(120x100)/120 = 60.00

Example 30. The following table gives the monthly average salary of a teacher and general price indices for a period of six years.

Year	:	1980	1981	1982	1983	1984	1985

Income (Rs.) : 360 420 500 550 600 640
General Price Index : 100 104 115 160 280 290

Find real average salary and construct real wage index based on 1980.

Solution :

Computation of Real Income and Real Income Indices

Year	Income (Rs.)	Price Index	Real Income	Real Income Index
1980	360	100	(360x100)/100 = 360.00	100
1981	420	104	(420x104)/100 = 403.80	1 12.2
1982	500	115	(500x115)/160 = 434.80	120.8
1983	550	160	(420x100)/160=343.80	95.5
1984	600	280	(420x100)/280=214.30	59.5
1985	640	290	(420x100)/290=220.70	61.3

Example 31. Compute the index of real wages from the following data using 1983 as base year.

Year :	1980	1981	1982	1983	1984	1985
Average Monthly Wages (Rs.) :	120	132	143	150	171	200
Price Index :	100	120	130	150	190	200

Solution : Computation of Real wage indices (Base 1983)

Year	Wages (Rs.)	Price Index	Real Wages	Real wages Index
				(Base = 1983)
1980	120	100	(120x100)/100 = 120	120
1981	132	120	(130x104)/120 = 110	110
1982	143	130	(143x115)/130 = 110	110
1983	150	150	(150x100)/150=100	100
1984	171	190	(171x100)/190= 90	90
1985	200	200	(200x100)/200= 100	100

6.3 Limitation of Index Numbers

Although index numbers are found to be very useful for measuring relative change in some phenomenon, they are not without limitation. These limitations are:

1. Index numbers in measure only approximate relative changes in two periods. They are capable of measuring changes in characteristics which can be quantified and vary with time.

2. Index numbers do not use complete data as only a limited number of representative items are included in their construction. Therefore, they do not reflect the true picture.
3. Selection of the base year is also a difficult task in the construction of index numbers as the selection of a 'normal year' is a subjective matter.
4. Determination of quality of the product is yet another important consideration in the construction of index numbers. However, it is a difficult task in modern times when qualities of different products undergo quick changes.
5. The index number is formed to serve only a specific purpose and, as such, its use is limited only to the phenomenon under study.
6. Index numbers are subjected to certain errors, namely-sampling errors, miscellaneous errors and incorrect classification of items. Sampling errors crop up in the selection of items, errors arising due to incomplete information, faulty price quotations while, lack of representative character of items come under miscellaneous errors. Classification of people into a specific class is an important problem in the construction of cost of living index numbers for a particular class of people. The classification may be faulty.

   In spite of these limitations, index number are regularly constructed and widely used for studying the related problems.

6.4 Check Your Progress

1. Factor reversal test is satisfied only by………………………………..
2. The method of construction of quantity index is same as that of ……………………..
3. In Laspeyer's method, weights are assigned to the commodities on the basis of …………………………… quantites.
4. In Paasche's method, weights are assigned to the commodities on the basis of …………………………… quantites.
5. In Fisher's method, weights are assigned to the commodities on the basis of both the …………………………… quantites.

6.5 Summary

When index numbers are constructed taking into consideration the importance of different commodities, they are called weighted index numbers. There are two methods of constructing weighted index numbers. First is weighted aggregative method which is further divided into different methods like Laspeyre's method, Paasche's method, Fisher' s method, Dorbish and Bowley method, Marshall- Edgeworth's method and Kelly's method; second is Weighted average of price relatives method. Quantity index numbers are designed to measure the changes in physical quantities of goods over a given period. It can also be simple or weighted. Various formulae can be used for the construction of index numbers. Prof. Fisher has given the following tests to select an appropriate formula: Time reversal test, Factor reversal test and circular test. Although index numbers are found to be very useful for measuring relative change in some phenomenon, they are not without limitation. In spite of these limitations, index number are regularly constructed and widely used for studying the related problems.

6.6 Keywords

Weighted Aggregative Method: In this method, commodities are assigned weights on the basis of the quantities purchased.

Weighted Average of Price Relatives Method: In this method, first of all, the price relatives for the current year are calculated on the basis of the base year prices of the commodities.

Time Reversal test: The test is that the formula for calculating an index number should be such that it will give the same ratio between one point of comparison and the other, no matter which of the two taken as base.

Circular test: Index number based on simple G.M., simple aggregative formula and weighted aggregative formula satisfy circular test.

6.7 Self- Assessment Test

Q1. Explain the methods of weighted index numbers.

Q2. From the following data, Price index for 1988 by using:

(a) Laspeyre's method

(b) Paasche's method

(c) Dorbish and Bowley's method

(d) Fisher's method

(e) Marshall-Edgeworth's method

Year	A	B	C	D

	Price	Quantity	Price	Quantity	Price	Quantity	Price	Quantity
1980	24	8	9	3	16	5	10	3
1988	30	10	10	4	20	8	9	4

(**Ans**: (a) 120.67 (b) 120.72 (c) 120.69 (d) 120.7 (e) 120.6)

Q3. Compute Quantity index number for the following data by (a) Simple Aggregative Method, (b) Average of Quantity relative method by using A.M.:

Commodity:	A	B	C	D	E	F
Production(1989)	20	30	10	25	40	50
Production(1999)	25	30	15	35	45	55

(**Ans.**: (a) 117.3 (b) 122.92)

Q4. Using suitable formula construct the price index number from the following data:

Commodity	**1990**		**1995**	
	Price	**Expenditure**	**Price**	**Expenditure**
A	1.0	60.00	1.25	62.50
B	1.50	37.50	2.50	50.00
C	2.00	20.00	3.00	30.00
D	12.00	36.00	18.00	72.00
E	0.10	4.00	0.15	9.00

Check whether it satisfies time reversal and factor reversal test.

6.8 Answers to Check Your Progress

1. Fisher’s Ideal Formula
2. Price Index
3. Base year
4. Current year
5. Base year as well as current year

6.9 References/Suggested Readings

1. Gupta, S. P.: Statistical Methods, Sultan Chand and Sons, New Delhi.
2. Levin, R. I. and David, S. R.: Statistics for Management, Prentice Hall, New Delhi.
3. Gupta, C. B.: Introduction to Statistical Methods.
4. Hooda, R. P.: Statistics for Business and Economics, Macmillan, New Delhi.

Subject: Business Statistics-1	
Course Code: 302	**Author :** Prof. Ved Paul
Lesson No. : 7	**Vetter:** Dr. B. S. Bodla
TIME SERIES ANALYSIS	

Structure

7.0 Learning Objectives

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

- Understand the concept time series analysis
- Understand the components of a time series
- Explain the methods of determination of trend
- Explain the methods of computational of seasonal indices

7.1 Introduction

A time series is an arrangement of statistical data in a chronological order, i.e., in accordance with its time of occurrence. It reflects the dynamic pace of movements of a phenomenon over a period of time. Most of the series relating to Economics, Business and Commerce are all time series spread over a long period of time. Accordingly, time series have an important and significant place in business and economics, and basically most of the statistical techniques for the analysis of time series data have been developed by economists. However, these techniques can also be applied for the study of behavior of any phenomenon collected chronologically over a period of time in any discipline relating to natural and social sciences, though not directly related to economics or business.

Time series analysis is a quantitative method we use to detect patterns of change in statistical information over regular intervals of time. We project these patterns to arrive at an estimate for the future. Thus, time series analysis helps us cope with uncertainty about the future. Moreover, the analysis of time series on major national aggregates such as population, national income, capital formation, etc., provides the most crucial information about the success and weakness of a growth strategy which may have been adopted in the past, and can serve as the basis of setting targets for the future.

7.1.1 Causes of variations in time series analysis

There are different causes of variation in time series analysis:-

1. *Seasonal effect (Seasonal Variation or Seasonal Fluctuations):* Many of the time series data exhibits a seasonal variation which is the annual period, such as sales and temperature readings. This type of variation is easy to understand and can be easily measured or removed from the data to give deseasonalized data. Seasonal Fluctuations describes any regular variation (fluctuation) with a period of less than one year for example cost of various types of fruits and vegetables, clothes, unemployment figures, average daily rainfall, increase in the sale of tea in winter, increase in the sale

of ice cream in summer, etc., all show seasonal variations. The changes which repeat themselves within a fixed period, are also called seasonal variations, for example, traffic on roads in morning and evening hours, Sales at festivals like EID, etc., increase in the number of passengers at weekend, etc. Seasonal variations are caused by climate, social customs, religious activities, etc.

2. *Other Cyclic Changes (Cyclical Variation or Cyclic Fluctuations):* Time series exhibits Cyclical Variations at a fixed period due to some other physical cause, such as daily variation in temperature. Cyclical variation is a non-seasonal component that varies in a recognizable cycle. Sometimes series exhibits oscillation which does not have a fixed period but is predictable to some extent. For example, economic data affected by business cycles with a period varying between about 5 and 7 years. In weekly or monthly data, the cyclical component may describe any regular variation (fluctuations) in time series data. The cyclical variation is periodic in nature and repeats itself like a business cycle, which has four phases (i) Peak (ii) Recession (iii) Trough/Depression (iv) Expansion.

3. *Trend (Secular Trend or Long Term Variation):* It is a longer-term change. Here we take into account the number of observations available and make a subjective assessment of what is long term. To understand the meaning of the long term, let for example climate variables sometimes exhibit cyclic variation over a very long time period such as 50 years. If one just had 20 years of data, this long term oscillation would appear to be a trend, but if several hundreds of years of data are available, then long term oscillations would be visible. These movements are systematic in nature where the movements are broad, steady, showing a slow rise or fall in the same direction. The trend may be linear or non-linear (curvilinear). Some examples of the secular trends are: Increase in prices, Increase in pollution, an increase in the need for wheat, increase in literacy rate, and decrease in deaths due to advances in science. Taking averages over a certain period is a simple way of detecting a trend in seasonal data. Change in averages with time is evidence of a trend in the given series, though there are more formal tests for detecting a trend in time series.

4. *Other Irregular Variation (Irregular Fluctuations)*: When trend and cyclical variations are removed from a set of time series data, the residual left, which may or may not be random. Various techniques for analyzing series of this type examine to see "if irregular variation may be explained in terms of probability models such as moving average or autoregressive models, i.e. we can see if any cyclical variation is still left in the residuals. These variations occur due to sudden causes are called residual variation (irregular variation or accidental or erratic fluctuations) and are unpredictable, for

example, a rise in prices of steel due to strike in the factory, accident due to failure of the break, flood, earth quick, war, etc.

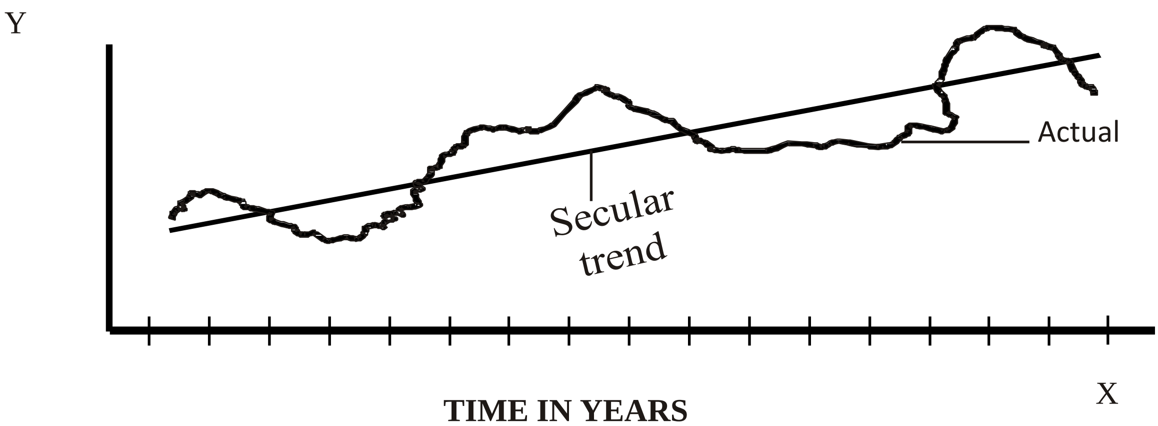
7.1.2 Components of a Time Series

As discussed above, a time series refers to any group of statistical information accumulated at regular intervals. Interestingly, if the values of a phenomenon are observed at different periods of time, the values so obtained will indicate appreciable variations or changes. These variations are due to the fact that the values of the phenomenon are affected not by a single factor but due to the cumulative effect of a multiplicity of factors pulling it up and down. For example, the price of a particular product depends on its demand, various competitive products in the market, raw materials and transportation expenses, investment and so on. There are four kinds of changes, or variations (known as components of time series), involved in time series analysis. They are:

1. Secular trend
2. Cyclical fluctuations
3. Seasonal variations
4. Irregular variations

Secular Trend

The general tendency of the time series data to increase or decrease or stagnate during a long period of time is called the secular trend or simple trend. In brief, the trend is the long-term movement of a time-series. The steady increase in the cost of living recorded by the Consumer Price Index is an example of secular trend. From year to year, the cost of living varies a great deal, but if examined over a long-term period, the trend toward a steady increase is observed. Figure 1 shows a trend in an increasing but fluctuating time series.



(FIG. 1)

There are four reasons why it is useful to study secular trend.

1. The study of trend enables us to describe a historical pattern. Many times a past trend is used to evaluate the success of a previous policy. For instance, an evaluation of the effectiveness of a recruiting programme by a university may be done on the basis of examination of its past enrolment trends.

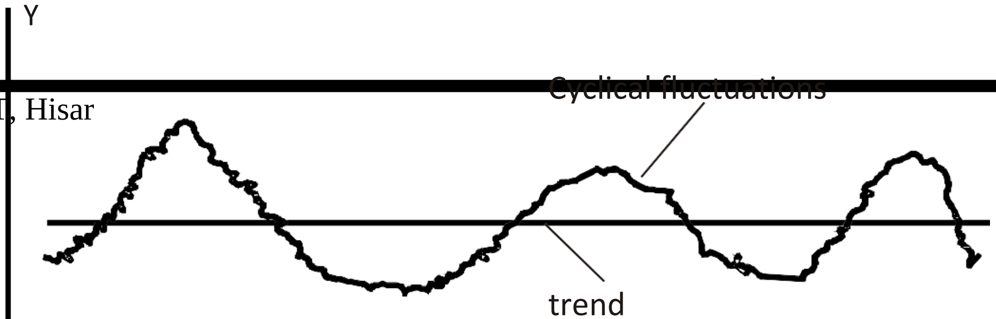
2. This helps in business forecasting and planning future operations. For example, if the time series data for a particular period regarding a particular phenomenon, say, growth rate of the India's population, is observed, we can estimate the population for some future time.

3. Trend makes it easier for us to study the other three components of the time series. This can be done by isolating trend values from the given time series.

4. Trend analysis enables us to compare two or more time series over different periods of time and draw important conclusions about them.

Cyclical Variations

The oscillatory movements in a time series with period of oscillation greater than one year are termed as cyclical variations. The most common example of cyclical fluctuation is the business cycle. These variations are the upswings and down wings in the time series that are observable over extended period of time. Neither the amplitude nor the frequency of occurrence of these cycles is uniform. Empirical studies based on the analysis of time series data on a large number of major economic aggregates for developed countries have shown that the length of time interval after which cycles occur ranges from 8 to 10 years. Figure 2 illustrates a typical pattern of cyclical movements. A knowledge of the cycle component enables a businessman to have an idea about the periodicity of the booms and depressions and accordingly he can take timely steps for maintaining stable market for his product.



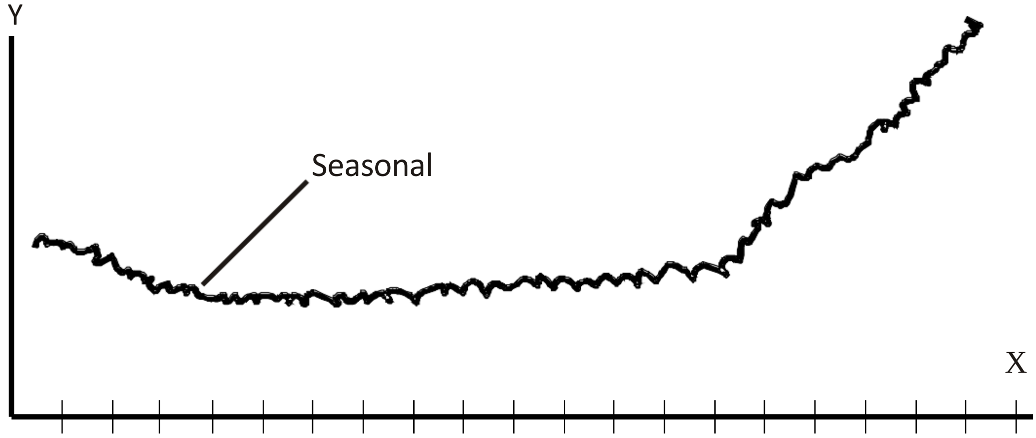
X

TIME IN YEARS
(FIG. 2)

Seasonal Variations

As we might expect from the name, seasonal variation involves patterns of change within a year that tend to be repeated from year to year. Some examples are the production of soft drinks, which is high during the summer and low during the winter; substantial increases in the number of flu cases every winter; woolen garments sales, which is high from October month to January and low during rest of the months.

In each of these examples, note that there are systematic causes of these fluctuations, such as the weather, holidays and government accounting procedures and so forth. These systematic causes occur regularly. Some other causes like national level festivals including Dussehra, Holi, and Diwali also bring a shift in sales of departmental stores. Since these variations are regular pattern, they are useful in forecasting the future. Fig. 3 illustrates a typical pattern of seasonal variations.



TIME IN MONTHS (FIG. 3)

Random or Irregular Variations

These variations are purely random and are the result of such unforeseen and unpredictable forces which operate in absolutely erratic and irregular manner. These powerful variations are caused by numerous non-recurring factors like floods, famines, wars, earthquakes, strikes, revolution, etc., which behave in a very unpredictable manner. The collapse of OPEC in 1986, the Iraqi situation in 1990 on gasoline prices in the United States, Security Prices fluctuations in India in March-April 1992 on account of SCAM are some examples of irregular variations.

7.1.3 Decomposition: additive and Multiplicative models

The following are the two models commonly used for the decomposition of a time series into its components:

1. $0_t = T_t + S_t + C_t + I_t$ .... Additive Model
2. $0_t = T_t \times S_t \times C_t \times I_t$.................... Multiplicative Model

Where 0_t is the time series value at time t, and T_t, S_t, C_t and l_t represent the trend, seasonal, cyclical and random variations at time t. In these models $S = S_t$, $C = C_t$ and $I = I_t$ are absolute quantities which can take positive and negative values so that

$\sum S = \sum S_t = 0,$ for any year

$\sum C = \sum C_t = 0,$ for any cycle and

$\sum I = \sum I_t = 0,$ in the long-term period.

The first model assumes that the economic time series is additive and is made up of the four components T, S, C and I. Here it is assumed that the four components are independent of one another. Independence is said to exist when the pattern of occurrence and the magnitude of movements in any particular component are not affected by the other components. As a concrete example, the production of beer has been increasing over last many years. This additive assumption implies that this steady increase in the production of beer has no effect on the seasonal variation of the production of beer. It also implies that the causes for the increase in the production of beer are different from the causes of the seasonal variation of beer. It may be noted that when the time series data are recorded against years, the seasonal component would vanish and in that case the additive model will take the form.

$0_t = T_t + C_t + R_t$

The second model - the multiplicative model, is used where it is assumed that the forces giving rise to the four types of variations are interdependent, so that the overall pattern of movements in the time series is the combined result of the interaction of all the forces operating on the time series. According to this assumption, the original magnitude of the time series is the product of its four components. The reason for using this model is that it allows convenient isolation of the components. If the decomposition of a time series is done by taking logarithms, the multiplication model will be expressed as

$$\text{Log } 0_t = \text{Log } T_t + \text{Log } C_t + \text{Log } S_t + \text{Log } R_t$$

Thus, we see that the four components of a time series relating to economic and business phenomenon conform to the multiplicative model. In practice, additive model is rarely used.

Decomposition of Time Series into Its Four Components

We have seen that there are two models of time series which can be used for decomposition of it. It could be observed from both of these models that decomposition of a time series requires estimation of its four components and then separating them from each other so as to be able to understand the pattern of variations in each component independently. We shall follow the Pearson's approach based on the multiplication model for decomposition of time series.

The first component to be estimated is the trend variations. Trend variations are estimated by fitting a trend line on the time series data. After estimating the trend variations, these are then separated from the time series is known as detrending. Detrending requires dividing both sides of Multiplicative model by the trend values T_t, so that

$$\frac{0_t}{T_t} = C_t \cdot S_t \cdot R_t$$

After isolating trend, we shall compute and separate the seasonal variations and the resulting multiplicative model would be expressed as

$$\frac{0_t}{} = C_t \cdot R_t$$

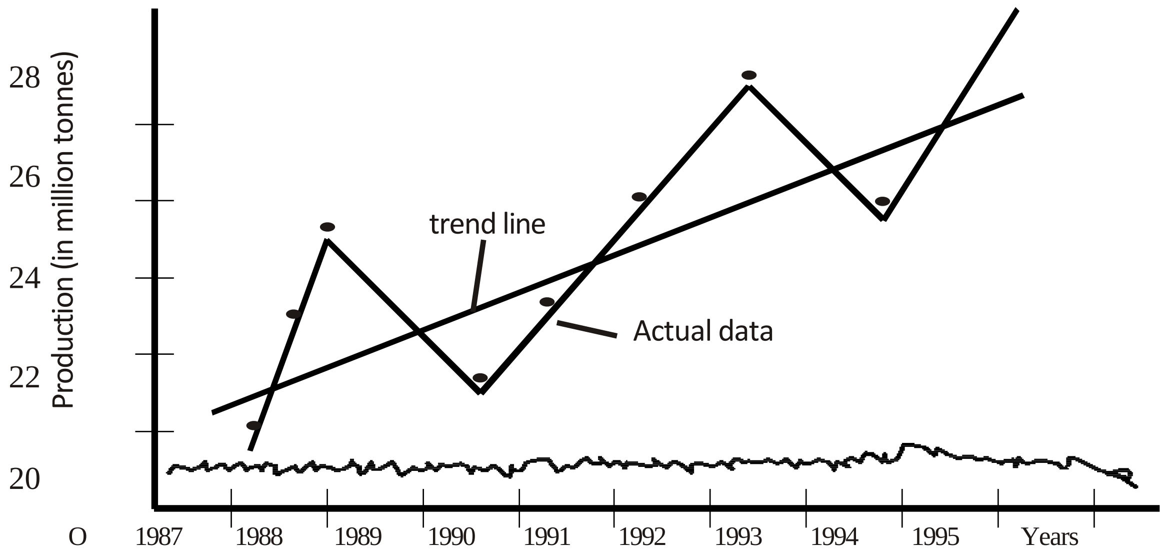
$T_t\ S_t$

So far we have discussed ways of separating the trend T and the seasonal variation S. The cyclical variations and irregular variations can be examined easily with reference to the pattern of their occurrence and amplitude. If we ignore the random variations then the values obtained in the following equation may be taken to reflect cyclical variations.

$$\frac{0_t}{T_t, S_t} = C_t . I_t$$

To arrive at better estimates of cyclical fluctuations, the irregular component

(I) should be eliminated from the $C_t . I_t$ value obtained in above equation. However, the extent to which their elimination is possible, they tend to become marginal in the process of deseasonalisation.



(FIG. 4)

7.2 Determination of Trend

Of the four components of a time series, secular trend represents the long- term direction of the series. There are various ways to describe the trend or fitting a straight line, such as the freehand curve method, the method of semi averages. The method of moving averages, and the method of least squares. In this lesson we shall discuss each of these methods in estimation of trend. The general formula for a straight line is Y = a+bx where x is called the independent variable, and Y is called the dependent variable is the Y-intercept of the straight line and b is the slope of the trend line.

The Free-hand Method

The simplest method of finding a trend line when given a set of time series data is the free hand method. According to this method first of all we shall plot the data on a graph and then, by observation, will fit a straight line through the plotted points in a way such that the straight line shows the trend of the time series.

Illustration-1: Fit a trend line to the following data by free hand method

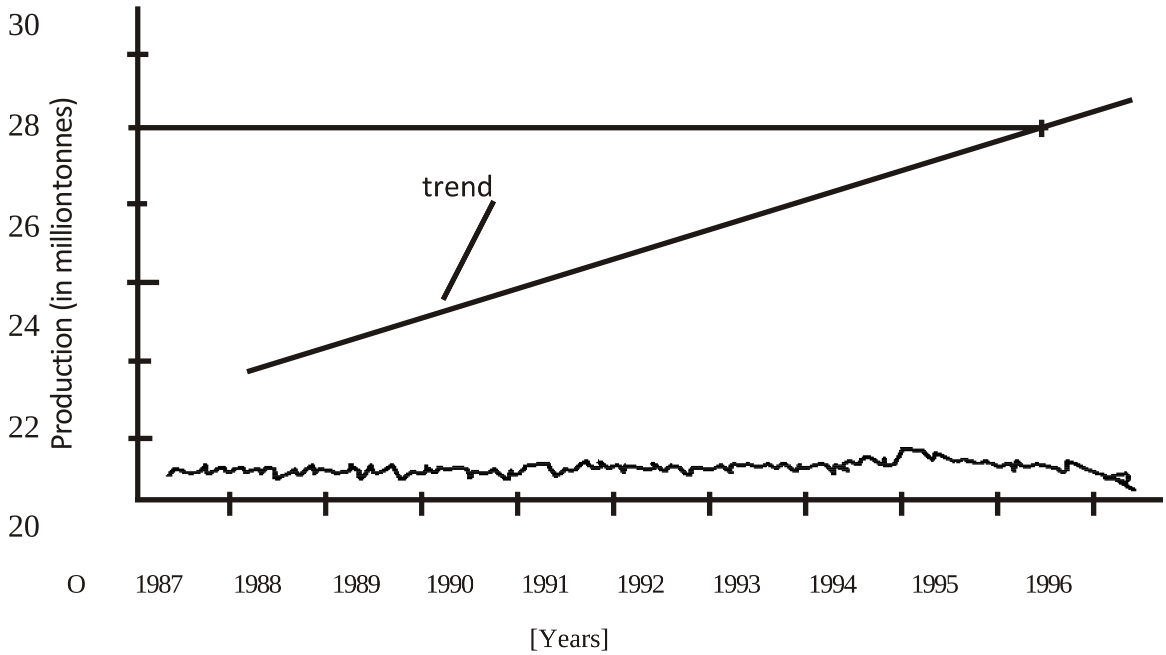
Year	***X***	***Production of Steel (in million tonnes)***	***Year***	***X***	***Production of Steel (in million tonnes)***
1987	-3	20	1992	2	25
1988	-2	22	1993	3	26
1989	-1	24	1994	4	25
1990	-0	21	1995	5	28
1991	1	23			

Solution:

From figure (4) it is obvious that this is not an accurate way of fitting a straight line or a curve to the data as it gives only rough idea regarding trend.

Finding the trend Equation: It requires selection of two points on the straight line. An important feature of time series is that the data are given in order of time. In illustration 1, it starts from 1987 and goes up to 1995 in one year time intervals. Let us start at 1987, and call it the 'origin', and designate it as zero. Then 1988 is 1, 1989 is 2 and so forth, as shown in illustration and also figure. In this way the origin may be placed at any year. If we let 1990 be the origin, then 1987 is -3, 1988 is -2, 1989 is -1, 1990 is 0, 1991 is 1, and so on.

Now, let us assume that the trend line goes through the points for 1987 and 1995 (illustration 1). It becomes a problem of finding the equation for the straight line going through the two points 1987 and 1995. The coordinates of the two points selected now become (-3, 20) and (5, 28). Substituting the values of these coordinates into the equation for a straight line, we find.



(FIG. 5)

20 = a + (-3b)

28 = a + 5b

Solving these two equations gives a = 17, b = 1. Thus the equation for the trend line is

$Y_C = 17 + 1X$ .... Y_C = Estimated value.

The interpretation of the equation is, when X = -3 (1987),

$Y_C = 17 + (1)(-3) = 17-3 = 14$

which indicates that the estimated production by the trend line is 14 million tonnes.

Method of Semi averages

This method divides the time series into two parts, finds the average of each part, and then fits a trend line through these averages. Note that in case number of years is even, middle year is left out for computing the semi averages.

Illustration-2: Using the statement of illustration-1 estimate the value for 1996 by applying method of semi averages.

Solution:

Here n = 9, and hence the two parts will be 1987 to 1990 and 1992 to 1995. As n is an odd number we will ignore the middle year of the series, i.e. 1991.

Year	***Production of Steel (in million tonnes)***	***4 years semi-total***	***Semi averges (A.M.)***
1987	20		
1988	22		
1989	24	87	21.75
1990	21		
1991	23		
1992	25		
1993	26		
1994	25	104	26.00
1995	28		

Here these m average 21.75 is to be plotted against the middle of the years 1988 and 1989 and value 26 against the middle of the years 1993 and 1994. The graph is shown in Fig (5). From the graph we see that the estimated value for 1996 is 27.7.

This method assumes the presence of linear trend which may not exist. Semi averages are affected by extreme values. This is a crude and simple way of fitting a trend line, but its simplicity is its advantage.

7.2.1 Moving Average Method

This is a very simple and flexible method of measuring trend. When a trend is to be determined by this method, the average value for a number of years (or months) is secured, and this average is taken as the normal or trend value for the unit of time falling at the middle of the period covered in the calculation of the average. The averaging process smoothens out fluctuations in the given data.

While applying moving average method, the choice of the length of the period for a moving average is necessary because this would determine the extent to which variations would be smoothened in the process of averaging. The period of moving average is to be decided in the light of the length of the cycle. These averages do not yield an equation which could be used for forecasting the values of a time series variable for the future.

Illustration 3: Estimate the trend values using the data given below by taking a three-yearly moving average.

Year	***Value***	***Year***	***Value***
1987	3	1992	11
1988	4	1993	09
1989	8	1994	10
1990	6	1995	14
1991	7		
Solution :			
Year	***Value***	***Three-yearly***	***Three yearly***
		moving-Total	***moving average***
1987	3		
1988	4	15	5
1989	8	18	6
1990	6	21	7
1991	7	24	8
1992	11	27	9
1993	9	30	10
1994	10	33	11
1995	14		

Note: If the moving average is an even period moving average, the moving total and moving average which are placed at the center of time span from which they are computed fall between two time periods.

The moving average method is applicable not only to trend lines but also to all kinds of data that show regular periodic fluctuations. We shall use it also to eliminate seasonal fluctuations.

7.2.2 The Method of Least Squares

The method of least square is the most widely used method of fitting a straight line to a series of data. Estimation of trend values by this method makes use of the general equation for estimating a straight line.

$Y_C = a+bx.$

The values of the two constants, a and b, in the Equation are obtained by solving simultaneously the two normal equations.

$\sum Y = na + b\sum x$

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

Where n represents number of years for which data are given. We can measure the variable X from any point of time in origin such as the first year. But the calculations are simplified when the mid-point in time is taken as the origin because in that case the negative values in the first half of the series balance out the positive values in the second half so that $\sum x = 0$. Since $\sum X = 0$ the above two normal equations would take the form

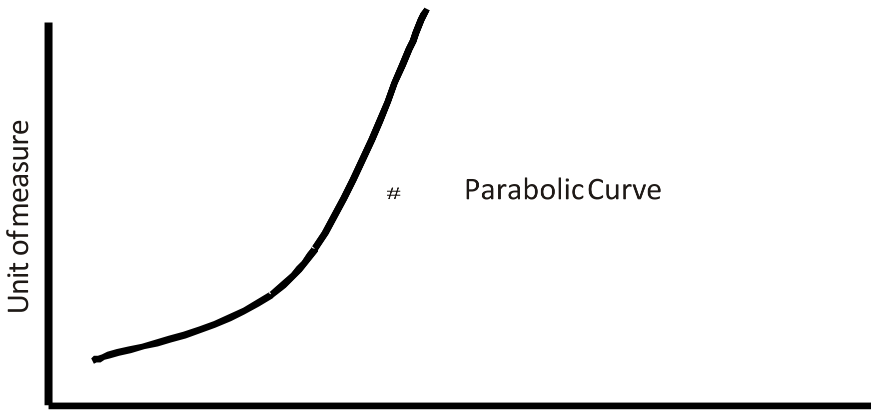
$\sum Y = na$..................................... (i)

$\sum XY = b\sum X^2$... (ii)

The values of a and b can now be determined easily. Since $\sum Y = na$, $a = \sum Y/N$

Since $\sum XY = b\sum X^2$, $b = \sum XY/\sum X^2$.

It should be noted that in case of odd number of years, when deviations are taken from the middle year. $\sum X$ would always be zero provided there is no gap in the data given. However, in case of even years also $\sum X$ will be zero if the X



TIME IN YEARS
(FIG. 6)

origin is placed midway between the two middle years.

Illustration : 4 Fit a trend line to the following data by the least squares method

Year	:	1991	1993	1995	1997	1999
Production (in '000 tons)	:	36	42	46	54	32

Estimate the production in 1996 and 2002.

Solution :

Let the trend line be given by the equation : Y = a+bx Where origin is at 1995.

Computation for Straight Line Trend

Year T	*Production* *(in '000 tons)* *(Y)*	*X=t-1989*	*XY*	*X^2*	*Trend values* *Y_c*
1991	36	- 4	-144	16	41.2
1993	42	-2	-84	4	41.6
1995	46	0	0	0	42.0
1997	54	2	108	4	42.4
1999	32	4	128	16	42.8
n=5	∑Y= 210	∑X = 0	∑XY = 8	$X^2 = 40$	

Since ∑X = 0, a = ∑ Y/N, b = $\sum XY/\sum X^2$

We have, ∑Y=210, N = 5, ∑XY = 8, $\sum X^2 = 40$

$$a = \frac{210}{5} = 42, \quad b = 8/40 = 0.2$$

Substituting values of a and b in the least square equation we get.

$$Y_c = 42 + 0.2X$$

By substituting X = -4, -2, 0, 2, 4, in the above equation we will obtain the estimated values for the years 1991, 1993, 1995, 1997 and 1999 respectively.

Thus

Y_{1991} = 42 + (0.2) (-4) = 42 - 0.8

= 41.2

Y_{1993} = 42 + (0.2) (-2) = 42 - 0.4 = 41.6

Y_{1999} = 42 + (0.2) (0) = 42.0

Y_{1996} = 42 + (0.2) (2) = 42.4

Y_{2002} = 42 + (0.2) (4) = 42.8

The estimated production in 1996 is obtained on taking X = t-1995 = 1996 - 1995 = 1

Y_{1996} = 42 + (0.2) (1) = 42.2

The estimated production in 2002 is obtained on taking X = t - 1995 = 2002 - 1995= 7

Hence Y_{2002} = 42 + (0.2) (7) = 43.4.

Second Degree Polynomial Trend

In the previous lesson, we have described the method of fitting a straight line to a time series. But many time series are best described by curves, not straight lines. The linear model does not adequately describe the change in the variable as time changes in case of non-linear trend. Most often a parabolic curve is used to overcome this problem. The parabolic curve is described mathematically by a second degree equation. A hypothetical parabolic curve is illustrated in Fig. (6). the general form for an estimated second-degree equation is:

$$Y_c = a+bX = CX^2$$

Where: Y_c is the estimated value of the dependent variable; a, b and c are numerical constants, and X represents the coded value of time variable.

It must be noted that if the third term CX^2 is introduced in Eq. , it will give a parabolic trend. In the above Eq. the value of c reveals whether the resultant second degree curve is concave or convex. Here the value of c also determines the extent to which the curve departs from linearity.

The derivation of the second-degree equation is beyond the scope of this lesson. However, we can determine the value of the numerical constants (a, b and c) from the following three equations:

$$\sum Y = an + c\sum X^2$$

$$\sum X^2 Y = a\sum X^2 + C\sum X^4$$

$$\sum XY \; b = \text{———}\sum X^2$$

After finding the values a, b and c by solving the above equations, we substitute these values in the equation for a second-degree parabola. A problem involving a parabolic trend is considered below in illustration.

Illustration 6 : Fit an equation of the form $Y = a+bX + cX^2$ to the data given below :

Years	1991	1992	1993	1994	1995
Consumption of wheat (in Qtls.)	25	28	33	39	46

Solution

Calculations for Second Degree Trend

Years	***Consumption (in Qtls.) Y***	***X***	***X²***	***X⁴***	***XY***	***X²Y***
1	25	- 2	4	16	-50	100
2	28	- 1	1	1	-28	28
3	33	0	0	0	0	00
4	39	1	1	1	39	39
5	46	2	4	15	92	184
n=5	$\sum Y = 171$	$\sum X = 0$	$\sum X^2 = 171$	$\sum Y^2 = 171$	$\sum XY = 171$	$\sum X^2Y=351$

We are to fit $Y_c = a+bX+cX^2$

The first step in fitting a second degree equation is translate the independent variable (time) into a coded time variable X. Not that the coded variable X is listed in one year intervals because there is an odd number of elements in our time series. Now find values of a, b and c by solving the three equations meant for the purpose.

Three equations are

$\sum Y = an + c\sum X^2$ ....(1)

$\sum X^2Y = a\sum X^2 + c\sum X^4$ ....(2)

$b = \dfrac{\sum XY}{\sum X^2}$ .... (3)

By substituting the given values in the above equations, we get : 171 = 5a + 10c 1 (a)

351 = 10a + 34c............................... 2 (a)

b = 5.3 .... 3 (a)

Now we shall find a and c by solving equations 1(a) and 2(a).

1. Multiply equation 1(a) by 2 and subtract equation 2(a) from equation 1(a)

342 = 10a + 20c

-351 = -10a - 34c

-9 = -14c .... (4)

From equation (4) we find c c = -9/-14 = 0.6

2. Substitute the value for c into equation 1(a).

171 = 5a+(10) (0.64)

171 = 5a + 6.4

164.6 = 5a

Lastly, we shall put these numerical values in the general equation as follows :

$Y_c = a + bx + cx^2$

$= 32.92 + 5.3x + 0.64x^2$

Exponential Trend:

$Y_c = ab^x$ .... (5.5)

Where : a and b are the two constants, and X represents the values assigned to time.
In general, the exponential trend is applicable, where growth in the time series data is nearly at a constant rate per unit of time (expressed in percentage). When the aggregate variable related to national product, population, or production in the country as a whole or in a region are given we normally use exponential trend.

Taking logarithm of both sides of Eq. (5), we get $\log Y = \log a + X \log b$ (6)

When plotted on a semi logarithmic graph, the curve gives a straight line. However, on an arithmetic chart the curve gives a non-linear trend.

To obtain the values of the two constants, a and b, we need to solve simultaneously the two normal equations :

$\sum \log Y = n \log a + \log b \sum X$

$\sum (X \log Y) = \log a \sum X + \log b \sum X^2$

When deviations are taken from middle year, i.e., $\sum X = 0$, the above equation takes the following form :

$\sum \log Y = n \log a$

$\sum \log y$ or $\log a = \frac{}{n}$

and $\sum (X, \log Y) = \log b \sum X^2$

$\sum \log b = \frac{\sum (X \log Y)}{}$

$\sum X^2$

The rate of growth implicit in a semi logarithmic trend is often of interest. It is derived by solving the equation for compound interest - log (1+r) = b_1

Here b_1 is the slope and r is the rate of growth.

Illustration7 : The sales of a company in lakhs of rupees for the years 1988 to 1994 are given below :

Years :	1988	1989	1990	1991	1992	1993	1994
Sales :	16	23	33	46	66	95	137

Estimates sales for the year 1995 using an equation of the form $Y = ab^x$, here X = years and Y = sales.

Solution: Fitting Equation of Form Y = ab

Year X	***Sales Y***	***X (Coded)***	***Log Y***	***X^2***	***X Log Y***
1988	16	-3	1.2041	9	-3.6123
1989	23	-2	1.3617	4	-2.7234
1990	33	-1	1.5185	1	-1.5185
1991	46	0	1.6627	0	0
1992	66	1	1.8195	1	1.8195
1993	95	2	1.9777	4	3.9554
1994	137	3	2.1367	9	6.4101
		$\sum X=0$	$\sum LogY=11.6819$	$\sum X^2=28$	$\sum XLog\ Y=4.3308$

$$\text{Log } a = \frac{\sum \text{LogY}}{} = \frac{11.6809}{} = 1.6687$$

n 7

$$\text{Log } b = \frac{\sum X \text{ Log}Y}{\sum X^2} = \frac{4.3308}{28} = 0.1547$$

We know, Log Y = Log a + X Log b

$\sum$ log Y = 1.6687 + 0.1547 X

For 1995, X would be +4. When X=4, Log Y will be - Log Y = 1.6687 + (0.1547) (4) = 2.2875

Y = Anti log 2.2875 = 193.86

Thus the estimated sales for the year 1995 is Rs. 193.86 lakhs.

7.3 Computation of Seasonal Indices

Seasonal variations are regular and periodic variations having a period of one year duration. Some of the examples which show seasonal variations are production of cold drinks, which are high during summer months and low during winter season. Sales of sarees in a cloth store which are high during festival season and low during other periods.

The reason for determining seasonal variations in a time series is to isolate it and to study its effect on the size of the variable in the index form which is usually referred as seasonal index.

Measurement of seasonal variations:

The study of seasonal variation has great importance for business enterprises to plan the production schedule in an efficient way so as to enable them to supply to the public demands according to seasons.

There are different devices to measure the seasonal variations. These are

- Method of simple averages.
- Ratio to trend method
- Ratio to moving average method
- Link relative method.

7.3.1 Method of simple averages

This is the simplest of all the methods of measuring seasonality. This method is based on the additive modal of the time series. That is the observed values of the series is expressed by $Y_t = \square T_t \square S_t \square C_t \square R_t$

and in this method we assume that the trend component and the cyclical component are absent.

Advantages and Disadvantages:

Method of simple average is easy and simple to execute.

This method is based on the basic assumption that the data do not contain any trend and cyclic components. Since most of the economic and business time series have trends and as such this method though simple is not of much practical utility.

Example: 1

Assuming that the trend is absent, determine if there is any seasonality in the data given below.

Year	*Ist Quarter*	*2nd Quarter*	*3rd Quarter*	*4th Quarter*
2004	3.7	4.1	3.3	3.5
2005	3.7	3.9	3.6	3.6
2006	4.0	4.1	3.3	3.1
2007	3.3	4.4	4.0	4.0

What are the seasonal indices for various quarters ? *(M. Com., M.K. Univ.)*

Solution. COMPUTATION OF SEASONAL INDICES

Year	*Ist Quarter*	*2nd Quarter*	*3rd Quarter*	*4th Quarter*
2004	3.7	4.1	3.3	3.5
2005	3.7	3.9	3.6	3.6
2006	4.0	4.1	3.3	3.1
2007	3.3	4.4	4.0	4.0
Total	14.7	16.5	14.2	14.2
Average	3.675	4.125	3.55	3.55
Seasonal Index	98.66	110.74	95.30	95.30

Notes for calculating seasonal index

The average of averages $= \frac{3.675 + 4.125 + 3.55 + 3.55}{4} = \frac{14.9}{4} = 3.725$

$$\text{Seasonal Index} = \frac{\text{Quarterly average}}{\text{General average}} \times 100$$

Seasonal Index for the first quarter $= \frac{3.675}{3.725} \times 100 = 98.66$

Seasonal Index for the second quarter $= \frac{4.125}{3.725} \times 100 = 110.74$

Seasonal Index for the third and fourth quarters $= \frac{3.55}{3.725} \times 100 = 95.30$

7.3.2 Ratio to trend method:

This method is an improvement over the simple averages method and this method assumes a multiplicative model i.e

$$Y_t \square T_t S_t C_t R_t$$

The measurement of seasonal indices by this method consists of the following steps.

- Obtain the trend values by the least square method by fitting a mathematical curve, either a straight line or second degree polynomial.
- Express the original data as the percentage of the trend values. Assuming the multiplicative model these percentages will contain the seasonal, cyclical and irregular components.
- The cyclical and irregular components are eliminated by averaging the percentages for different months (quarters) if the data are In monthly (quarterly), thus leaving us with indices of seasonal variations.
- Finally these indices obtained in step(3) are adjusted to a total of 1200 for monthly and 400 for quarterly data by multiplying them through out by a constant K which is given by K = 1200/Total No. of Indicies; for monthly and K = 400/Total No. of Indicies; for Quarterly.

Advantages:

- It is easy to compute and easy to understand.
- Compared with the method of monthly averages this method is certainly a more logical procedure for measuring seasonal variations.
- It has an advantage over the ratio to moving average method that in this method we obtain ratio to trend values for each period for which data are available where as it is not possible in ratio to moving average method.

Disadvantages:

- The main defect of the ratio to trend method is that if there are cyclical swings in the series, the trend whether a straight line or a curve can never follow the actual data as closely as a 12- monthly moving average does. So a seasonal index computed by the ratio to moving average method may be less biased than the one calculated by the ratio to trend method.

Example 1: Calculate seasonal indices by Ratio to trend method from the following data.

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2003	30	40	36	34
2004	34	52	50	44
2005	40	58	54	48
2006	54	76	68	62
2007	80	92	86	82

Solution. For determining seasonal variation by ratio-to-trend method, first we will determine the trend for yearly data and then convert it to quarterly data.

CALCULATING TREND BY METHOD OF LEAST SQUARES

Year	Yearly totals	Yearly average Y	Deviations from mid-year X	XY	X^2	Trend values
2003	140	35	− 2	− 70	4	32
2004	180	45	− 1	− 45	1	44
2005	200	50	0	0	0	56
2006	260	65	+ 1	+ 65	1	68
2007	340	85	+ 2	+ 170	4	80
$N = 5$		$\Sigma Y = 280$		$\Sigma XY = 120$	$\Sigma X^2 = 10$	

The equation of the straight line trend is $Y = a + bX$.

$$a = \frac{\Sigma Y}{N} = \frac{280}{5} = 56 \qquad b = \frac{\Sigma XY}{\Sigma X^2} = \frac{120}{10} = 12$$

Quarterly increment $= \frac{12}{4} = 3.$

Calculation of Quarterly Trend Values. Consider 2003, trend value for the middle quarter, *i.e.*, half of 2nd and half of 3rd is 32. Quarterly increment is 3. So the trend value of 2nd quarter is $32 - \frac{3}{2}$, *i.e.*, 30.5 and for 3rd quarter is $32 + \frac{3}{2}$, *i.e.*, 33.5. Trend value for the 1st quarter is 30.5 − 3, *i.e.*, 27.5 and of 4th quarter is 33.5 + 3, *i.e.*, 36.5. We thus get quarterly trend values as shown below :

TREND VALUES

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2003	27.5	30.5	33.5	36.5
2004	39.5	42.5	45.5	48.5
2005	51.5	54.5	57.5	60.5
2006	63.5	66.5	69.5	72.5
2007	75.5	78.5	81.5	84.5

The given values are expressed as percentage of the corresponding trend values.

Thus for 1st Qtr. of 2003, the percentage shall be $(30/27.5) \times 100 = 109.09$, for 2nd Qtr. $(40/30.5) \times 100 = 131.15$, etc.

GIVEN QUARTERLY VALUES AS % OF TREND VALUES

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2003	109.09	131.15	107.46	93.15
2004	86.08	122.35	109.89	90.72
2005	77.67	106.42	93.91	79.34
2006	85.04	114.29	97.84	85.52
2007	105.96	117.20	105.52	97.04
Total	463.84	591.41	514.62	445.77
Average	92.77	118.28	102.92	89.15
S.I. Adjusted	92.05	117.36	102.12	88.46

Total of averages = 92.77 + 118.28 + 102.92 + 89.15 = 403.12.

Since the total is more than 400 an adjustment is made by multiplying each average by $\frac{400}{403.12}$ and final indices are obtained.

7.3.3 Ratio to moving average method:

The ratio to moving average method is also known as percentage of moving average method and is the most widely used method of measuring seasonal variations. The steps necessary for determining seasonal variations by this method are

- Calculate the centered 12-monthly moving average (or 4-quarterly moving average) of the given data. These moving averages values will eliminate S and I leaving us T and C components.

- Express the original data as percentages of the centered moving average values.
- The seasonal indices are now obtained by eliminating the irregular or random components by averaging these percentages using A.M or median.
- The sum of these indices will not in general be equal to 1200 (for monthly) or 400 (for quarterly). Finally the adjustment is done to make the sum of the indices to a total of 1200 for monthly and 400 for quarterly data by multiplying them throughout by a constant K which is given by

  K =1200/Total of the indices……………for monthly

  K =400/Total of the indices……………for Quarterly

Advantages:

- Of all the methods of measuring seasonal variations, the ratio to moving average method is the most satisfactory, flexible and widely used method.
- The fluctuations of indices based on ratio to moving average method is less than based on other methods.

Disadvantages:

This method does not completely utilize the data. For example in case of 12-monthly moving average seasonal indices cannot be obtained for the first 6 months and last 6 months.

Illustration 24. Calculate seasonal indices by the ratio to moving average method, from the following data :

Year	*1st Quarter*	*2nd Quarter*	*3rd Quarter*	*4th Quarter*
2005	68	62	61	63
2006	65	58	66	61
2007	68	63	63	67

Solution. CALCULATION OF SEASONAL INDICES BY 'RATIO TO MOVING AVERAGE' METHOD

Year	*Quarter*	*Given figures*	*4-figure moving totals*	*2-figure moving totals*	*4-figure moving average*	*Given figure as % of moving average*
2005	I	68				
	II	62				
			254			
	III	61		505	63.186	96.54
			251			
	IV	63		498	62.260	101.19
			247			
2006	I	65		499	62.375	104.21
			252			
	II	58		502	62.750	92.43
			250			
	III	66		503	62.875	104.97
			253			
	IV	61		511	63.875	95.50
			258			
2007	I	68		513	64.125	106.04
			255			
	II	63		516	64.500	97.67
			261			
	III	63				
	IV	67				

CALCULATION OF SEASONAL INDEX

Year	*Percentage to Moving Average*			
	1st Quarter	*2nd Quarter*	*3rd Quarter*	*4th Quarter*
2005	—	—	96.63	101.20
2006	104.21	92.43	104.97	95.50
2007	106.04	97.67	—	—
Total	210.25	190.10	201.60	196.70
Average	105.125	95.05	100.80	98.35
Seasonal Index	105.30	95.21	100.97	98.52

$$\text{Arithmetic average of averages} = \frac{399.32}{4} = 99.83$$

By expressing each quarterly average as percentage of 99.83, we will obtain seasonal indices.

$$\text{Seasonal index of 1st Quarter} = \frac{105.125}{99.83} \times 100 = 105.30$$

$$\text{Seasonal index of 2nd Quarter} = \frac{95.05}{99.83} \times 100 = 95.21$$

$$\text{Seasonal index of 3rd Quarter} = \frac{100.80}{99.83} \times 100 = 100.97$$

$$\text{Seasonal index of 4th Quarter} = \frac{98.35}{99.83} \times 100 = 98.52$$

7.3.4 Link relative method:

This method is slightly more complicated than other methods. This method is also known as Pearson‟s method. This method consists in the following steps.

- The link relatives for each period are calculated by using the below formula Link relative for any period = Current periods figure/ Pr evious periods figure ×100
- Calculate the average of the link relatives for each period for all the years using mean or median.
- Convert the average link relatives into chain relatives on the basis of the first season.

Chain relative for any period can be obtained by = Avg link relative for that period / Chain relative of the previous period × 100 the chain relative for the first period is assumed to be 100.

- Now the adjusted chain relatives are calculated by subtracting correction factor „kd" from (k+1)th chain relative respectively.
- Where k = 1,2,…….11 for monthly and k = 1,2,3 for quarterly data.

And d = $\frac{1}{N}\{Chain\ relative\ for\ the\ first\ period - 100\}$

where N denotes the number of periods i.e. N = 12 for monthly N = 4 for quarterly

- Finally calculate the average of the corrected chain relatives and convert the corrected chain relatives as the percentages of this average. These percentages are seasonal indices calculated by the link relative method.

Advantages:

- As compared to the method of moving average the link relative method uses data more.

Disadvantages:

- The link relative method needs extensive calculations compared to other methods and is not as simple as the method of moving average.
- The average of link relatives contains both trend and cyclical components and these components are eliminated by applying correction.

Deseasonalisation

When the seasonal component is removed from the original data, the reduced data are free from seasonal variations and is called deseasonalised data. That is, under a multiplicative model

$$\frac{T \times S \times C \times I}{S} = T \times C \times I.$$

Deseasonalised data being free from the seasonal impact manifest only average value of data.

Seasonal adjustment can be made by dividing the original data by the seasonal index.

$$\text{That is, Deseasonalised data} = \frac{\text{Original data}}{\text{Seasonal index}} \times 100$$

where an adjustment-multiplier 100 is necessary because the seasonal indices are usually given in percentages.

In case of additive model

$$Y_t = T + S + C + I,$$

$$\text{Deseasonalised data} = \text{Original data} - \frac{\text{Seasonal index}}{100}$$

$$= Y_t - \frac{\text{Seasonal index}}{100}.$$

Uses and limitations of seasonal indices

Seasonal indices are indices of seasonal variation and provide a quantitative measure of typical seasonal behavior in the form of seasonal fluctuations.

7.4 Check Your Progress

There are some activities to check your progress. Fill in the blanks:

(a) The analysis of time series consists in decomposition of a time series into its basic……………..

(b) Semi-Average method is affected by……..............................

(C) Least square method can be used to fit ……………………………………or………………………or…………………………….

(D) Method of simple averages is used in those situations where trend is assumed to be ………..……………………………. In the data.

(E) Ratio to trend method assumes a………………………………….. model.

7.5 Summary

A time series is an arrangement of statistical data in a chronological order, i.e., in accordance with its time of occurrence. There are four kinds of changes, or variations (known as components of time series), involved in time series analysis. They are: Secular trend, cyclical fluctuations, Seasonal variations and Irregular variations. There are two models of time series which can be used for decomposition of it. It could be observed from both of these models that decomposition of a time series requires estimation of its four components and then separating them from each other so as to be able to understand the pattern of variations in each component independently. There are various ways to describe the trend or fitting a straight line, such as the freehand curve method, the method of semi averages, the method of moving averages, and the method of least squares. Seasonal variations are regular and periodic variations having a period of one year duration. Some of the examples which show seasonal variations are production of cold drinks, which are high during summer months and low during winter season. There are different devices to measure the seasonal variations. These are Method of simple averages, Ratio to trend method, Ratio to moving average method and Link relative method.

7.6 Keywords

Deseasonalised: When the seasonal component is removed from the original data, the reduced data are free from seasonal variations and is called deseasonalised data.

Secular Trend: The general tendency of the time series data to increase or decrease or stagnate during a long period of time is called the secular trend or simple trend.

Cyclic Variations: The oscillatory movements in a time series with period of oscillation greater than one year are termed as cyclical variations.

Seasonal Variation: As we might expect from the name, seasonal variation involves patterns of change within a year that tend to be repeated from year to year.

Random or irregular variations: These variations are purely random and are the result of such unforeseen and unpredictable forces which operate in absolutely erratic and irregular manner.

The method of least square: It is the most widely used method of fitting a straight line to a series of data.

Moving Average Method: When a trend is to be determined by this method, the average value for a number of years (or months) is secured, and this average is taken as the normal or trend value for the unit of time falling at the middle of the period covered in the calculation of the average. The averaging process smoothens out fluctuations in the given data.

Method of simple averages: This is the simplest of all the methods of measuring seasonality.

This method is based on the additive modal of the time series.

Ratio to trend method: This method is an improvement over the simple averages method and this method assumes a multiplicative model.

Ratio to moving average method: The ratio to moving average method is also known as percentage of moving average method and is the most widely used method of measuring seasonal variations.

7.7 Self-Assessment Test

Q1. Distinguish between seasonal variations, and cyclical fluctuations. How would you measure secular trend in any given data?

Q2. Describe the method of link relatives for calculating the seasonal variation indices.

Q3. How would you determine seasonal variation in the absence of trend?

Q4. Briefly describe the relative merits and demerits of ratio to trend and ratio to moving average method.

Q5. What do you understand by cyclical fluctuations in time series?

Q6. What do you understand by random fluctuation in time series?

Q7. Explain the term „Business cycle" and point out the necessity of its study in time series analysis.

Q8. Calculate seasonal variation for the following data of sale in thousands Rs. of a firm by the Ratio to trend method.

Q9. What is a time-series? What are its main components?

Q10.What do you mean by decomposition of a time series?

Q11. Distinguish between additive and multiplicative model in the analysis of time series.

Q12. What is 'Secular Trend'?Discuss the various ways of estimating the trend value.

Q13. Fit a trend line from the following data by using semi-average method:

Year :	1993	1994	1995	1996	1997	1998
Profits :	100	120	140	150	130	200
(in '000 Rs.)						

Answer: Joining the points (1994, 120) and (1997, 160) we get the trend line.

7.8 Answers to Check Your Progress

(a) Components

(b) Extreme values

(c) Straight line trend or parabolic trend or exponential trend

(d) Absent

(e) Multiplicative

7.9 References/Suggested Readings

1. Gupta, S. P.: Statistical Methods, Sultan Chand and Sons, New Delhi.
2. Levin, R. I. and David, S. R.: Statistics for Management, Prentice Hall, New Delhi.
3. Gupta, C. B.: Introduction to Statistical Methods.