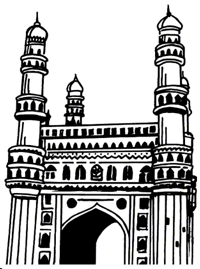
Rahul's ✔
Topper's Voice

R19
New Syllabus



J N T U (H)
M B A

I Year I Semester

Latest **2022** Edition

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

- **Study Manual**
- **FAQ's and Important Questions**
- **Short Question & Answers**
- **Internal Assessment**
- **Solved Previous Question Papers**
- **Solved Model Papers**

Price
` 199-00

- by -

WELL EXPERIENCED LECTURER

Rahul Publications ™
Hyderabad. Ph : 66550071, 9391018098

JNTU(H) MBA

I Year I Semester

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Price `. 199 -00

Sole Distributors : ☎ : **66550071, Cell : 9391018098**

VASU BOOK CENTRE

Shop No. 2, Beside Gokul Chat, Koti, Hyderabad.

Maternity Hospital Opp. Lane, Narayan Naik Complex, Koti, Hyderabad.

Near Andhra Bank, Subway, Sultan Bazar, Koti, Hyderabad -195.

CONTENTS

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

STUDY MANUAL

SOLVED MODEL PAPERS

SOLVED PREVIOUS QUESTION PAPERS

SYLLABUS

UNIT - I

Introduction to Research - Types of Research, Research Process-Conceptualization of variables and Measurement – Types and measurement of variables – Reliability and validity in measurement of variables- sources of error in measurement- Ethics in business research.

UNIT - II

Research design - Research Problem- purpose of Research design, Types of Research Design- Experimental research design, Research Design for cross sectional, longitudinal studies, Research design for action research – Characteristics of the good research design. Data Collection Methods & Tools: Types of Data, Sources and Instruments for data, Guidelines for questionnaire, Sampling and its application.

UNIT - III

(a) **Tabulation** of Univariate, Bivariate and multivariate data, Data classification and tabulation, Diagrammatic and graphical representation of data. One dimensional, two dimensional and three-dimensional diagrams and graphs

(b) **Small Sample Tests -** t-Distribution-properties and applications, testing for one and two means, paired t-test.

UNIT - IV

(a) **Analysis of Variance -** One Way and Two-Way ANOVA (with and without Interaction). Chi-Square distribution: Test for a specified Population variance, Test for Goodness of fit, Test for Independence of Attributes.

(b) **Correlation Analysis -** correlation, limits for coefficient of Correlation, Karl Pearson's coefficient of correlation, Spearman's Rank correlation, Linear and Multiple regression analysis, Discriminant analysis, Exploratory Factor Analysis. Introduction to Structural Equation Modeling, Cluster Analysis and Conjoint Analysis.

UNIT - V

Time Series Analysis and Report Writing:

(a) Components, Models of Time Series–Additive, Multiplicative and Mixed models; Trend Analysis- Free hand curve, Semi averages, moving averages, Least Square methods and Index numbers – introduction, Characteristics and uses of index numbers, types of index numbers, unweighted price indexes, weighted price indexes, Tests of adequacy and consumer price indexes.

(b) Importance of Report writing, Types of Research Reports, Report Preparation and presentation, Report structure, Report formulation, Guides for effective documentation, Research Briefings. Referencing styles and citation in Business Management Research.

Contents

UNIT - I

UNIT - II

Topic		Page No.

UNIT - III

UNIT - IV

Topic **Page No.**

UNIT - V

Frequently Asked & Important Questions

UNIT - I

1. Explain the classification of research.

Ans : **(Jan.-20, Imp.)**

Refer Unit-I, Q.No. 4.

---

2. Outline the process of research.

Ans : **(Jan.-20, Imp.)**

Refer Unit-I, Q.No. 5.

---

3. Define measurement. What are the characteristics of measurement?

Ans : **(Aug.-21, Nov.-20, Imp.)**

Refer Unit-I, Q.No. 9.

---

4. What are the different types of measure-ment scales with examples?

Ans : **(Aug.-21, Nov.-20)**

Refer Unit-I, Q.No. 13.

---

5. What are the various sources of errors in measurement.

Ans : **(Imp.)**

Refer Unit-I, Q.No. 16.

---

6. What do you understand by ethics in business research?

Ans : **(Imp.)**

Refer Unit-I, Q.No. 17.

UNIT - II

1. Define Research Design. What are the components of Research Design?

Ans : **(Jan.-20, Imp.)**

Refer Unit-II, Q.No. 3.

---

2. What are the different types of research designs.

Ans : **(Nov.-20, Jan.-20, Imp.)**

Refer Unit-II, Q.No. 7.

3. What are the Characteristics of the Good Research Design?

Ans : **(Jan.-20, Imp.)**

Refer Unit-II, Q.No. 10.

4. Explain the various methods of collecting primary data.

Ans : **(Jan.-20, Imp.)**

Refer Unit-II, Q.No. 14.

5. What are the difference between Primary and Secondary Data ?

Ans : **(Imp.)**

Refer Unit-II, Q.No. 20.

6. Define Questionnaire. Explain different types of Questionnaire.

Ans : **(Imp.)**

Refer Unit-II, Q.No. 21.

7. What are the Guidelines / Precautions for Preparation of Questionnaire ?

Ans : **(Imp.)**

Refer Unit-II, Q.No. 24.

8. Explain different types of Sampling Methods.

Ans : **(Imp.)**

Refer Unit-II, Q.No. 31.

9. What are the Applications of Sampling?

Ans : **(Imp.)**

Refer Unit-II, Q.No. 33.

UNIT - III

1. What is tabulation ? Explain the signifi- cance of tabulation.

Ans : **(Jan.-20, Imp.)**

Refer Unit-III, Q.No. 1.

2. Explain different parts of a table.

Ans : **(Jan.-20, Imp.)**

Refer Unit-III, Q.No. 2.

3. Explain different types of classification?

Ans : **(Imp.)**

Refer Unit-III, Q.No. 8.

4. What do you understand bivariate / multivariate tabulation.

Ans : **(Nov.-21, Imp.)**

Refer Unit-III, Q.No. 6.

5. Define Small sample tests and t-distribution.

Ans : **(Nov.-21, Jan.-20)**

Refer Unit-III, Q.No. 21.

6. Explain the properties and applications of t-distribution.

Ans : **(Nov.-21, Imp.)**

Refer Unit-III, Q.No. 22.

7. Explain briefly about Paired t-test.

Ans : **(Jan.-20)**

Refer Unit-III, Q.No. 25.

UNIT - IV

1. What is ANOVA? What are its assump-tions and applications?

Ans : **(Jan.-20, Imp.)**

Refer Unit-IV, Q.No. 1.

2. Explain briefly about One Way ANOVA.

Ans : **(Jan.-20, Imp.)**

Refer Unit-IV, Q.No. 2.

3. Explain briefly about two way ANOVA with and without interation?

Ans : **(Jan.-20, Imp.)**

Refer Unit-IV, Q.No. 3.

4. Explain briefly about test for goodness of fit?

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 7.

5. What are the different types of corelations.

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 10.

6. Define scatter diagram. Explain merits and demerits of scatter diagram.

Ans : **(Nov.-20, Imp.)**

Refer Unit-IV, Q.No. 14.

7. What is Karl Pearson's Coefficient of Correlation? Explain properties of Coefficient of Correlation.

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 17.

8. Define the term Regression. Explain the utility of Regression Analysis.

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 22.

9. What is Discriminant Analysis? Explain the objectives of discriminant analysis?

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 31.

10. Define Exploratory Factor Analysis. What are the objectives of Exploratory Factor Analysis?

Ans : **(Jan.-20, Imp.)**

Refer Unit-IV, Q.No. 36.

11. Define Cluster Analysis. Explain the features of Cluster Analysis.

Ans : **(Aug.-21, Imp.)**

Refer Unit-IV, Q.No. 45.

UNIT - V

1. Explain the Utility of Time Series Analysis.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 2.

2. What are the Components of Time Series ?

Ans : **(Nov.-20, Imp.)**

Refer Unit-V, Q.No. 3.

3. What do you understand by index numbers ?

Ans : **(Aug.-21, Imp.)**

Refer Unit-V, Q.No. 12.

4. Explain the Characteristics of Index Numbers.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 13.

5. Explain the uses of index numbers.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 14.

6. Explain the various methods of constructing index numbers.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 17.

7. Define Report, Report Writing and Research Report. State the importance of Report Writing.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 26.

8. Explain the process of report preparation and presentation.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 28.

9. Describe the structure of an effective report.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 30.

10. Discuss the referencing styles and citation in business management research.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 34.

UNIT I	**Introduction to Research -** Types of Research, Research Process-Conceptualization of variables and Measurement – Types and measurement of variables – Reliability and validity in measurement of variables- sources of error in measurement- Ethics in business research.

1.1 Introduction to Research

Q1. What do you understand research?

Ans : **(Jan.-20)**

The term research refers to an activity where a study is carried out in a particular discipline or a subject of particular interest, especially to discover new facts or information in the discipline or the concerned subject of interest area. This study may focus on the behaviour of the phe-nomenon or phenomena, their relationship, or interrelationship. It may be mentioned that we observe and relate to phenomenon or phenomena both in physical and social sciences and can be termed as the primary objective of research. While doing this we build belief on their behaviour and their relationship.

Building beliefs on the behaviour of phenomena in the physical world or in society/ economy can be done through common sense or with a scientific approach. However, the ways of building beliefs by a common man with his common sense and the scientist with a scientific approach are different. Consider the following events that are generally observed by people in various fields.

- An object falling to earth when dropped from a height.
- Demand and prices of normal commodities are inversely related.
- Authoritarian managers are not successful.
- Women are more religious than men.
- Brand preference and economic status are positively related.
- People from a caste or ethnic group are enterprising.
- Investment in company X is profitable.
- Sales and advertising expenditures are positively related.

Meaning of Research

Research is a scientific inquiry aimed at learning new facts, testing ideas, etc. It is the systematic collection, analysis and interpretation of data to generate new knowledge and answer a certain question or solve a problem.

- Research is the systematic process of collecting and analyzing information to increase our understanding of the phenomenon under study. It is the function of the research to contribute to the understanding of the phenomenon and to communicate that understanding to others.

Definitions of Research

According to Clifford Woody "Research comprises of defining and redefining problems, formulating hypothesis or suggested solutions, making deductions and reaching conclusions; and a last carefully testing the conclusions to determine whether they fit the formulating hypothesis".

According to Redman and Mory "Research is a systematized effort to gain new knowledge".

According to Advanced Learner's Dictionary "Research is a careful investigation or inquiry especially through search for new facts in any branch of knowledge.

Q2. What are the Characteristics of Research?

Ans :

Research is defined as the scientific investigation of phenomena which includes collection, presentation, analysis and interpretation of facts that lines an individual's speculation with reality.

Characteristics of Research

1. **Empirical:** Research is based on direct experience or observation by the researcher.
2. **Logical:** Research is based on valid procedures and principles.
3. **Cyclical:** Research is a cyclical process because it starts with a problem and ends with a problem.
4. **Analytical:** Research utilizes proven analytical procedures in gathering the data, whether historical, descriptive, experimental and case study.
5. **Critical:** Research exhibits careful and precise judgment.
6. **Methodical:** Research is conducted in a methodical manner without bias using systematic method and procedures.
7. **Replicability:** The research design and procedures are replicated or repeated to enable the researcher to arrive at valid and conclusive results.

Q3. Explain the Scope of Research.

Ans :

1. Decision-Making Tool

Research is useful for taking marketing management decisions. It provides necessary information and data in analysed and processed forms for making marketing decisions. With advanced technology, higher production functions and an increasing marketing complex, market research has become an indispensable tool for taking appropriate decisions.

2. Management Plannings

Research is used for management planning. It deals with marketing opportunities, i.e., those opportunities which are viable to be exploited by management. Thus, management can assess the resources that will be useful for the business.

3. Problem-Solving

Starting from problem identification to formulation of alternative solutions, and evaluating the alternatives in every area of management, is the problem-solving action of research. Problem-solving research focuses on the short-range and long-range decisions that must be taken with respect to the elements of the marketing mix, viz., product, price, place and promotion. It can help managements bring about prompt adjustment and innovations in the above areas of management.

4. Control Technique

Research is used as a control technique of management to find out the weaknesses and shortcoming of the management decisions to re-orient the planning and performance techniques.

5. Large-Scale Production

Research helps large-scale production by providing suitable decisions to be undertaken by the producers to exploit the existing production resources to meet the growing markets. The resources of production and market potentials are properly assessed by research.

6. Complex Market

The advancement of science and technology and the standard of living of consumers necessitate closer touch with the growing markets. The size and specialisation within the business unit and the intervention of numerous middlemen between the manufacturer and customers created a wide communication gap. The widening gap requires marketing research to fill up the communication gap between the consumer and the producer.

7. Pattern of Consumption

The pattern of consumption is to be assessed by the management. The study of buyers' behaviour, attitudes and capacity to purchase is very important in research. The purchasing power of a consumer depends upon his disposable personal income. Thus, the total purchasing power of a country or geographical area can be assessed by the disposable income of the place. The research reveals all the factors which influence the pattern of consumption.

8. Market Complex

The marketing activities are influenced by several internal and external environments. Internal environments include price, promotion and production and place (distribution), whereas the external environments include economic, sociological, political, legal and government motives.

9. Suitable Marketing Operations

Marketing operations decide production functions, and marketing operations can be better decided by the findings of marketing research. Marketing functions are concerned with the maximising of profit, and production functions have to minimise the cost of production. The blending of these two functions will give a higher margin to the company.

10. Pricing

Pricing is not arbitrary for follow-up action of competitors. It has to be judiciously fixed which is done effectively with the study of various marketing variables. The pricing objectives, market share, payment procedures, market demand, elasticity of demand, competitor's attitudes, price, volume relationship and changes in various market variables are studied by undertaking marketing research to frame suitable pricing policies.

11. Marketing Strategy

Marketing management has to lay down appropriate marketing strategies to meet competition to pursue growth in the market and to attain organisational objectives.

12. Distribution

Research helps the members of the channel of distribution to formulate suitable policies and programmes to solve their problem.

13. Sales Promotion

Research can decide suitable media of sales promotion after a study of the various channels of promotion. The costs and benefits of advertising, personal selling and wide publicity should be studied to decide the most appropriate media of sales promotion.

14. Helps Discharge Managerial Functions

Research helps the management to discharge its managerial functions of planning, forecasting, coordinating, motivating, controlling and evaluation. Research being a fact-finding process, significantly influences business declensions.

15. Production

Research is a must in the production area. Research and development helps invention, developing new products and modifying existing products. Research plays a significant role in the identification of a new project.

1.2 Types of Research

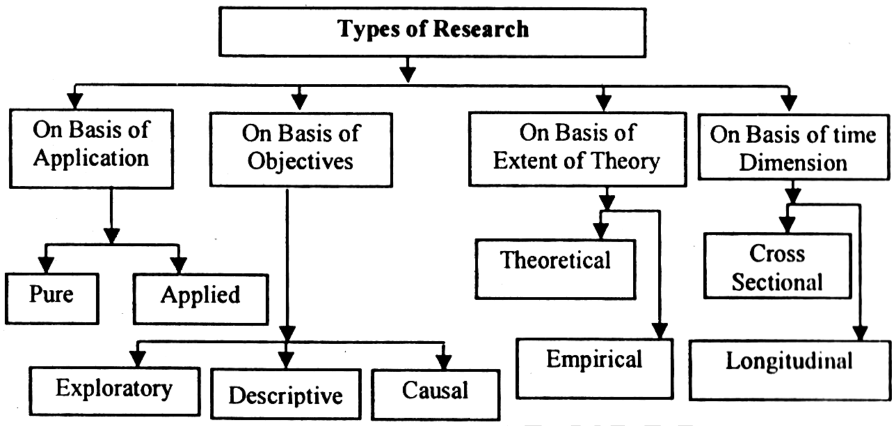
Q4. What are the different types of Research?

(or)

Explain the classification of research.

Ans : **(Jan.-20, Imp.)**

There are several ways to classify the research which are as follows :



1. **On the Basis of Application :** On the basis of application research is of two types :

i) **Pure/Basic/Fundamental Research:** This research is also called basic research, is concerned with quest for knowledge more about the phenomenon without concern for its practical use and also with developing and testing hypotheses and theories. Pure research takes place to explore a particular concept, or issue, without regards for a specific problem, and may be carry out to simple gain a better understanding of the overall concept. It is said, there is nothing as practical as a good theory. For example the development of a model of a coaching behaviour. Such research in itself has no immediate value beyond contributing to an area of intellectual inquiry. It is undertaken for increase in knowledge.

ii) **Applied Research :** This research is concern with search for ways of using scientific knowledge to solve practical problems. It focuses on analysing and solving social and real life problem. The findings become basis of framing programme and policies, based on principles of pure research. According to Horton and Hunt, this research is an investigation for ways of using scientific knowledge to solve practical problems. Because this research is generally conducted on large-scale basis, it is expensive. As such, it is often conducted with the support of some financing agency like government, public corporation. World Bank, UNICEF, UGC, ICSSR, etc. Many a time, this type of research is conducted on interdisciplinary basis also.

2. **On the Basis of Objective :** On the basis of fundamental objective research are designed in following ways:

i) **Exploratory Research :** When the purpose of research is to gain familiarity with a phenomenon or acquire new insights into it in order to formulate a more precise problem or develop hypothesis, the exploratory studies (also known as formulative research) come in handy. If the theory happens to be too general or too specific, a hypothesis cannot be formulated. Therefore a need for an exploratory research is felt to gain experience that will be helpful in formulating relevant hypothesis for more definite investigation.

ii) **Descriptive Research :** Descriptive research, also known as statistical research, describes data and characteristics about the population or phenomenon being studied. Descriptive research answers the questions who, what, where, when, and how. The description is used for frequencies, averages, and other statistical calculations. Often the best approach, prior to writing descriptive research, is to conduct a survey investigation. In short descriptive research deals with everything that can be counted and studied. But there are always restrictions to that. The research must have an impact to the lives of the people around the researcher.

iii) **Causal Experimental Research :** Although experimental research is primarily possible in areas of physical sciences, with the help of hypothesis, may also be carried out in social sciences if such research enables us to quantify the findings, to apply the statistical and mathematical tools and to measure the results thus quantified. It is also classified under conclusive research. For example, to investigate "what is the effect of mineral content of water on plant growth".

Causal research is used to obtain evidence of cause-and-effect (causal) relationships. In casual studies it is typical to have an expectation of the relationship to be explained, such as prediction about the influence of price, packaging, advertising, and the like on sales. Thus, researcher must be knowledgeable about the research Subject.

3. **On the Basis of Extent of Theory :** On the basis of extent of theory research are of two types:

i) **Theoretical Research :** Theoretical research generally uses the findings from existing works to develop new ideas through analysing existing theory and explanations. These new ideas are not tested through collecting evidence in the form of primary data. Theoretical research is held to be a classical way of adding something of value to the body of knowledge.

ii) **Empirical Research :** It is a way of gaining knowledge by means of direct observation or experience. Empirical evidence (the record of one's direct observations or experiences) can be analysed quantitatively or qualitatively. It is based on observation and experi-ence more than upon theory and abstraction and can be both qualitative and quantitative. Health research mainly follows the empirical approach. An empirical research in- Dives these five stages.

4. **On the Basis of Time Dimension :** On the basis of time dimension Research can be of two types:

i) **Cross-Sectional Research :** In cross-sectional research, researchers observe at one point in time. Cross-sectional research is usually the simplest and least costly alternative A cross-sectional design provides a snapshot of the variables included in the study, at one particular point in time. It may reveal how those variables are represented in a cross-section of a population. Cross-sectional designs generally use survey techniques to gather data, e.g., the U.S. Census. Cross-sectional research can be exploratory, descriptive, or explanatory but it is most consistent with a descriptive approach to research.

ii) **Longitudinal Research :** Researchers using longitudinal research examine features of people or other units at more than one time. It is usually more complex and costly than cross-sectional research, but it is also "more powerful, especially when researchers seek answers to questions about social change. Descriptive and explanatory researchers use longitudinal approaches. It will now consider three types of longitudinal research which are as follows :

a) **Time-Series Research :** A time series design collects data on the same variable at regular intervals (weeks, months, years, etc.) in the form of aggregate measures of a population. Measure-ments are taken on each variable over two or more distinct time periods. This allows the researcher to measure change in variables over time. For example, the Consumer Price Index (CPI), the FBI Uniform Crime Rate, unemployment rates, poverty rates, etc. Time series designs are useful for establishing a baseline measure, describing changes over time, Keeping track of trends, and Forecasting future (short-term) trends.

b) **Panel Study :** It is a powerful type of longitudinal research. It is more difficult to conduct than time-series research. In a p;inel study, the researcher observes exactly the same people, group, or organisation across time periods. Participants, who are examined over repeated time points may be affected by having previously completed the measure being used. (This is known as sensitisation). Panel research is formidable to conduct and very costly. Tracking people over time is often difficult because some people die or cannot be located.

c) **Cohort Study :** Cohort analysis is "explicitly macroanalytic", which means researchers examine the category as a whole for important features. In cohort study, the individuals examined over time may not be the same but they should be representative of a particular group (or cohort) of individuals who have shared a common experience.

1.3 Research Process

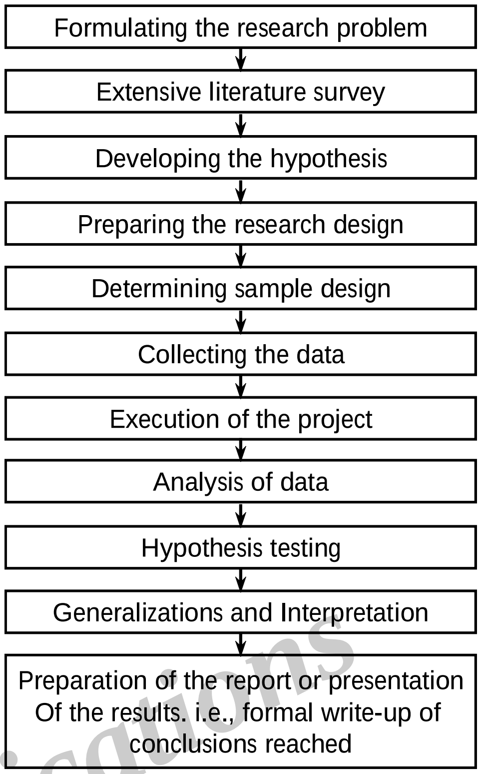
Q5. Explain the steps involved in research process.

(or)

Outline the process of research.

Ans : **(Jan.-20, Imp.)**

The following are the steps involved in research process.



A brief description of the above stated steps will be helpful.

1. Formulating the Research Problem

There are two types of research problems, viz., those which relate to states of nature and those which relate to relationships between variables. At the very outset the researcher must single out the problem he wants to study, i.e., he must decide the general area of interest or aspect of a subject-matter that he would like to inquire into. Initially the problem may be stated in a broad general way and then the ambiguities, if any, relating to the problem be resolved. Then, the feasibility of a particular solution has to be considered before a working formulation of the problem can be set up.

The formulation of a general topic into a specific research problem, thus, constitutes the first step in a scientific enquiry. Essentially two steps are involved in formulating the research problem, viz., understanding the problem thoroughly, and rephrasing the same into meaningful terms from an analytical point of view.

The best way of understanding the problem is to discuss it with one's own colleagues or with those having some expertise in the matter. In an academic institution the researcher can seek the help from a guide who is usually an experienced man and has several research problems in mind.

Often, the guide puts forth the problem in general terms and it is up to the researcher to narrow it down and phrase the problem in operational terms.

In private business units or in governmental organisations, the problem is usually earmarked by the administrative agencies with whom the researcher can discuss as to how the problem originally came about and what considerations are involved in its possible solutions.

2. Extensive Literature Survey

Once the problem is formulated, a brief summary of it should be written down. It is compulsory for a research worker writing a thesis for a Ph.D. degree to write a synopsis of the topic and submit it to the necessary Committee or the Research Board for approval. At this juncture the researcher should undertake extensive literature survey connected with the problem.

For this purpose, the abstracting and indexing journals and published or unpublished bibliographies are the first place to go to. Academic journals, conference proceedings, government reports, books etc., must be tapped depending on the nature of the problem. In this process, it should be remembered that one source will lead to another. A good library will be a great help to the researcher at this stage.

3. Development of Working Hypotheses

After extensive literature survey, researcher should state in clear terms the working hypothesis or hypotheses. Working hypothesis is tentative assumption made in order to draw out and test its logical or empirical consequences. As such the manner in which research hypotheses are developed is particularly important since they provide the focal point for research.

They also affect the manner in which tests must be conducted in the analysis of data and indirectly the quality of data which is required for the analysis. In most types of research, the development of working hypothesis plays an important role. Hypothesis should be very specific and limited to the piece of research in hand because it has to be tested.

The role of the hypothesis is to guide the researcher by delimiting the area of research and to keep him on the right track. It sharpens his thinking and focuses attention on the more important facets of the problem. It also indicates the type of data required and the type of methods of data analysis to be used.

How does one go about developing working hypotheses? The answer is by using the following approach:

(a) Discussions with colleagues and experts about the problem, its origin and the objectives in seeking a solution;

(b) Examination of data and records, if available, concerning the problem for possible trends, peculiarities and other clues;

(c) Review of similar studies in the area or of the studies on similar problems; and

(d) Exploratory personal investigation which involves original field interviews on a limited scale with interested parties and individuals with a view to secure greater insight into the practical aspects of the problem.

4. Preparing the Research Design

The research problem having been formulated in clear cut terms, the researcher will be required to prepare a research design, i.e., he will have to state the cdnceptual structure within which research would be conducted.

The preparation of such a design facilitates research to be as efficient as possible yielding maximal information. In other words, the function of research design is to provide for the collection of relevant evidence with minimal expenditure of effort, time and money. But how all these can be achieved depends mainly on the research purpose.

Research purposes may be grouped into four categories, viz.,

(i) Exploration,

(ii) Description,

(iii) Diagnosis, and

(iv) Experimentation.

A flexible research design which provides opportunity for considering many different aspects of a problem is considered appropriate if the purpose of the research study is that of exploration. But when the purpose happens to be an accurate description of a situation or of an association between variables, the suitable design will be one that minimises bias and maximises the reliability of the data collected and analysed.

There are several research designs, such as, experimental and non-experimental hypothesis testing. Experimental designs can be either informal designs (such as before-and-after without control, after-only with control, before-and-after with control) or formal designs (such as completely randomized design, randomized block design, Latin square design, simple and complex factorial designs), out of which the researcher must select one for his own project.

The preparation of the research design, appropriate for a particular research problem, involves % usually the consideration of the following:

(i) The means of obtaining the information;

(ii) The availability and skills of the researcher and his staff (if any);

(iii) Explanation of the way in which selected means of obtaining information will be organised and the reasoning leading to the selection;

(iv) The time available for research; and

(v) The cost factor relating to research, i.e., the finance available for the purpose.

5. Determining Sample Design

All the items under consideration in any field of inquiry constitute a 'universe' or 'population'. A complete enumeration of all the items in the 'population' is known as a census inquiry. It can be presumed that in such an inquiry when all the items are covered no element of chance is left and highest accuracy is obtained. But in practice this may not be true.

Even the slightest element of bias in such an inquiry will get larger and larger as the number of observations increases. Moreover, there is no way of checking the element of bias or its extent except through a resurvey or use of sample checks. Besides, this type of inquiry involves a great deal of time, money and energy. Not only this, census inquiry is not possible in practice under many circumstances. For instance, blood testing is done only on sample basis.

Hence, quite often we select only a few items from the universe for our study purposes. The items so selected constitute what is technically called a sample.

The researcher must decide the way of selecting a sample or what is popularly known as the sample design. In other words, a sample design is a definite plan determined before any data are actually collected for obtaining a sample from a given population. Thus, the plan to select 12 of a city's 200 drugstores in a certain way constitutes a sample design. Samples can be either probability samples or non-probability samples.

With probability samples each element has a known probability of being included in the sample but the non-probability samples do not allow the researcher to determine this probability. Probability samples are those based on simple random sampling, systematic sampling, stratified sampling, cluster/area sampling whereas non-probability samples are those based on convenience sampling, judgement sampling and quota sampling techniques.

Collecting the data: In dealing with any real life problem it is often found that data at hand are inadequate, and hence, it becomes necessary to collect data that are appropriate. There are several ways of collecting the appropriate data which differ considerably in context of money costs, time and other resources at the disposal of the researcher.

Primary data can be collected either through experiment or through survey. If the researcher conducts an experiment, he observes some quantitative measurements, or the data, with the help of which he examines the truth contained in his hypothesis. But in the case of a survey, data can be collected by any one or more of the following ways:

(i) By observation

This method implies the collection of information by way of investigator's own observation, without interviewing the respondents. The information obtained relates to what is currently happening and is not complicated by either the past behaviour or future intentions or attitudes of respondents. This method is no doubt an expensive method and the information provided by this method is also very limited. As such this method is not suitable in inquiries where large samples are concerned.

(ii) Through personal interview

The investigator follows a rigid procedure and seeks answers to a set of pre-conceived questions through personal interviews. This method of collecting data is usually carried out in a structured way where output depends upon the ability of the interviewer to a large extent.

(iii) Through telephone interviews

This method of collecting information involves contacting the respondents on telephone itself. This is not a very widely used method but it plays an important role in industrial surveys in developed regions, particularly, when the survey has to be accomplished in a very limited time.

(iv) By mailing of questionnaires

The researcher and the respondents do come in contact with each other if this method of survey is adopted. Questionnaires are mailed to the respondents with a request to return after completing the same. It is the most extensively used method in various economic and business surveys. Before applying this method, usually a Pilot Study for testing the questionnaire is conduced which reveals the weaknesses, if any, of the questionnaire. Questionnaire to be used must be prepared very carefully so that it may prove to be effective in collecting the relevant information.

(v) Through schedules

Under this method the enumerators are appointed and given training. They are provided with schedules containing relevant questions. These enumerators go to respondents with these schedules. Data are collected by filling up the schedules by enumerators on the basis of replies given by respondents. Much depends upon the capability of enumerators so far as this method is concerned. Some occasional field checks on the work of the enumerators may ensure sincere work.

6. Execution of the project

Execution of the project is a very important step in the research process. If the execution of the project proceeds on correct lines, the data to be collected would be adequate and dependable. The researcher should see that the project is executed in a systematic manner and in time. If the survey is to be conducted by means of structured questionnaires, data can be readily machine-processed. In such a situation, questions as well as the possible answers may be coded. If the data are to be collected through interviewers, arrangements should be made for proper selection and training of the interviewers. The training may be given with the help of instruction manuals which explain clearly the job of the interviewers at each step. Occasional field checks should be made to ensure that the interviewers are doing their assigned job sincerely and efficiently.

A careful watch should be kept for unanticipated factors in order to keep the survey as much realistic as possible. This, in other words, means that steps should be taken to ensure that the survey is under statistical control so that the collected information is in accordance with the pre-defined standard of accuracy. If some of the respondents do not cooperate, some suitable methods should be designed to tackle this problem.

One method of dealing with the non-response problem is to make a list of the non-respondents and take a small sub-sample of them, and then with the help of experts vigorous efforts can be made for securing response

7. Analysis of data

After the data have been collected, the researcher turns to the task of analysing them. The analysis of data requires a number of closely related operations such as establishment of categories, the application of these categories to raw data through coding, tabulation and then drawing statistical inferences.

The unwieldy data should necessarily be condensed into a few manageable groups and tables for further analysis. Thus, researcher should classify the raw data into some purposeful and usable categories. Coding operation is usually done at this stage through which the categories of data are transformed into symbols that may be tabulated and counted.

Editing is the procedure that improves the quality of the data for coding. With coding the stage is ready for tabulation. Tabulation is a part of the technical procedure wherein the classified data are put in the form of tables. The mechanical devices can be made use of at this juncture. A great deal of data, specially in large inquiries, is tabulated by computers. Computers not only save rune but also make it possible to study large number of variables affecting a problem simultaneously.

8. Hypothesis-testing

After analysing the data as stated above, the researcher is in a position to test the hypotheses, if any, he had formulated earlier.

Various tests, such as Chi square test, f-test, F-test, have been developed by statisticians for the purpose. The hypotheses may be tested through the use of one or more of such tests, depending upon the nature and object of research inquiry. Hypothesis-testing will result in either accepting the hypothesis or in rejecting it. If the researcher had no hypotheses to start with, generalisations established on the basis of data may be stated as hypotheses to be tested by subsequent researches in times to come.

9. Generalisations and interpretation

If a hypothesis is tested and upheld several times, it may be possible for the researcher to arrive at generalisation, i.e., to build a theory. As a matter of fact, the real value of research lies in its ability to arrive at certain generalisations. If the researcher had no hypothesis to start with, he might seek to explain his findings on the basis of some theory. It is known as interpretation. The process of interpretation may quite often trigger off new questions which in turn may lead to further researches.

10. Preparation of the report or the thesis

Finally, the researcher has to prepare the report of what has been done by him. Writing of report must be done with great care keeping in view the following:

11. The layout of the report should be as follows:

(i) The preliminary pages;

(ii) The main text, and

(iii) The end matter

In its preliminary pages the report should carry title and date followed by acknowledgements and foreword. Then there should be a table of contents followed by a list of tables and list of graphs and charts, if any, given in the report.

The main text of the report should have the following parts:

(a) Introduction: It should contain a clear statement of the objective of the research and an explanation of the methodology adopted in accomplishing the research. The scope of the study along with various limitations should as well be stated in this part.

(b) **Summary of findings:** After introduction there would appear a statement of findings and recommendations in non-technical language. If the findings are extensive, they should be summarised.

(c) **Main report:** The main body of the report should be presented in logical sequence and broken-down into readily identifiable sections.

(d) **Conclusion:** Towards the end of the main text, researcher should again put down the results of his research clearly and precisely. In fact, it is the final summing up.

Q6. Explain the criteria of good research.

Ans :

1. The purpose of the research should be clearly defined and common concepts be used.
2. The research procedure used should be described in sufficient detail to permit another researcher to repeat the research for further advancement, keeping the continuity of what has already been attained.
3. The procedural design of the research should be carefully planned to yield results that are as objective as possible.
4. The researcher should report with complete frankness, flaws in procedural design and estimate their effects upon the findings.
5. The analysis of data should be sufficiently adequate to reveal its significance and the methods of analysis used should be appropriate. The validity and reliability of the data should be checked carefully.
6. Conclusions should be confined to those justified by the data of the research and limited to those for which the data provide an adequate basis.
7. Greater confidence in research is warranted if the researcher is experienced, has a good reputation in research and is a person of integrity.

1.4 Conceptualization of Variables

Q7. Define the term variable. Explain various types of variables.

Ans :

Variable

A variable can be simply defined as any factor that varies. It is a certain aspect of an experimental condition that can vary or take on different characteristics in different conditions. The values of variables can change at different times for the same person or object or at the same time for different persons or objects. They are vital because they help to focus on specific events out of the many that are associated with the phenomenon. They connect the empirical world with the theoretical world. Health status, age, attitudes, memory, production units, motivation etc., are some of the examples of variables.

Types of Variables

In research study, variables are classified into five types,

1. Independent Variables

The variables which are supposed to be responsible for bringing about change(s) in a phenomenon are called independent variables. They refers to the presumed "cause" variations in which result in variations in the status of the dependent variable. They are also known as predictor or explanatory or regressor variables. In experimental research, independent variable is the one that is systematically controlled and manipulated by the researcher to estimate its overall effect on final outcome or dependent variable, whereas in non-experimental research, it can be described as the variable that has some logical influence on a dependent variable.

2. Dependent Variables

These variables are outcome of the changes brought about by the introduction of independent variable. They represent the response observed or measured by the researcher after making changes to the independent variable. Dependent variables refers to the status of the "effect" (outcome) in which the experimenter is interested. Any change in the independent variable produces a change in the dependent variable, which is observed and recorded. Researchers are primarily interested in dependent variables. This is because, by analyzing dependent variable, the researcher can find solutions to the problem. They are also known as response/ outcome/experimental variables. To establish that a change in independent variable produces a change independent variable, the following conditions must be met.

- A change in independent variable must be linked with a change in dependent variable.
- The independent variable (causal factor) should occur before the dependent variable (effect).
- No other possible factors should be responsible for the change in dependent variable i.e., all other possible causal factors must be controlled.
- A logical reasoning is required to explain why the independent variable affects the dependent variable.

3. Moderating Variable

Moderating variable is a variable which is not an independent variable but constitutes strong circumstances which may effect the relationship of independent and dependent variables. For example, relationship between training and performance in an organization. Training may be considered as an independent variable where performance as a dependent variable. 'Willingness to learn' is a variable which is neither independent nor dependent, but have a determined effect on the relationship between training and performance. Because employees who are more willing to learn will perceive training effectively as a result their performance will be improved. If an employee is not willing to learn but they are trained, then there will be no improvement in the performance. Thus, willingness to lear becomes moderating variable.

4. Extraneous Variables (Confounding Variables)

Several other factors operating during the experimental stages may practically influence the changes in the dependent variable. Therefore, all such factors/variables that might affect the results of the experiment must be controlled. These other variables are termed as extraneous or confounding variables. Therefore, confounding variables can be defined as the variables that can affect the outcome of a study in different ways that are not intended by the experimenter. These variables possess two properties as stated below,

(i) They are associated with the indepen-dent variable in the sense that individuals who differ for the independent variable also differ for the confounding variable.

(ii) They also affect the (dependent) variable. Thus, both independent and confounding variables cause a change in the dependent variable. However, there is no way to establish how much change is due to independent variable and how much is due to confounding variable. They mask the effects of independent variable on the dependant variable. There are two types of extraneous variables. They are,

(a) Situational Variables: These variables are the experimental conditions that might influence the response variables. Egs: Light, temperature, noise etc. They must be controlled so that they are constant for each phenomenon.

(b) Participant/Subject Variables: These refer to the ways in which one participant/subject differs form the other and how this influences the experimental results. E.gs: Mood, intelligence, concentration etc.

5. Intervening Variable

Independent and dependent variables which are involved in a study might affect the dependent variable with another variable/factor. It is not possible to observe or measure exactly.

Let us consider an example of a retail store. The sales of a retail store may increase if it gives discounts (Eg: 2%, 3%, 4%, 5%, ...). Suppose, retail store introduced a scheme under which monthly prizes are given to customers who are selected randomly. As a result the sales of the retail store may increase where it cannot be measured than this would be considered as intervening variable.

6. Control of Confounding Variables

Confounding variable, if left uncontrolled produces a confounding (confusing) effect on results of the research study. The researcher must identify these variables early in the study and make efforts to control them. Different strategies for controlling confounding variables are enlisted below,

(a) Eliminate Confounding Variables from Study: This is one of the best ways to control confounding variables.

Example

If a researcher is interested in studying the effects of exercise on physical activity in Indian adults then the sample should include only Indian adults. Children and other enthnicities should not be included in the study.

(b) Random Selection: Random sample selection should be done whenever feasible. One of the best ways to randomize the sample is to ensure that everyone in the sample population has an equal chance of being selected in the study.

(c) Random Assignment: When random selection is not feasible, then random assignment of sample to treatment groups must be done to reduce the effect of confounding.

(d) Statistical Control: When eliminating or randomizing the potential confounding variables is not possible, the next best thing is to statistically control them. Using an analysis of covariance (ANOVA) is an example of statistical control. This procedure helps to remove the effects of confounding variable from the analysis.

(e) Matching: The researcher can match subjects in the experimental and control groups on the potential confounding variable. However it is a time consuming and difficult process and hence is less preferred.

Q8. Compare and contrast between independent variable and dependent variable.

Ans :

The differences between independent and dependent variables are follows,

Independent variable	Dependent variable
1. It is responsible for bringing about change(s) in a phenomenon.	1. It is the outcome of the change (s) brought about by the introduction of the independent variable.
2. It is controlled and manipulated by the researcher.	2. It is a measurable variable which is difficult to manipulate.
3. It takes the form of experiment stimulus having two attributes i.e., present or absent.	3. It has attributes which are indirect, direct or through constructs.
4. It is a "casual" variable.	4. It is considered as a "caused" variable
5. It appears, disappears, or varies as the researcher introduces, eliminates or varies it.	5. It is observed and measured to determine the effect of independent variable.
6. It is the presumed cause, changes in which leads to changes in the status of dependent variable.	6. It is the presumed effect (response) in which the researcher is interested.
7. It is known as predictor, explanatory, regressor, controlled or input variable.	7. It is known as response, regressand, measured, explained, experimental output variable.

1.5 Conceptualization of Measurement

1.5.1 Types and Measurement of Variables

Q9. Define measurement. What are the characteristics of measurement?

Ans : **(Nov.-20, Imp.)**

Measurement is the process of assigning numbers or labels to objects, persons, states, or events in accordance with specific rules to represent quantities or qualities of attributes. Measurement then, is a procedure used to assign numbers that reflect the amount of an attribute possessed by an event, person, or object. Note that the event, person, or object is not being measured, but rather its attributes are. A researcher, for example, does not measure a consumer but measures attitude, income, brand loyalty, age, and other relevant factors.

Definition of Measurement

According to stevens measurement is the assignment of numerals to objects or events according to rules.

"**According to Campbell** measurement is the assignment of numbers to represent properties."

Characteristics of Measurement Techni- ques

If the measures are good the data will be good. Data become informative only after they get communicated. The following are six desirable characteristics in a measuring instrument whether it is to be used to collect data :

a) **Relevance:** Relevance is the primary contributor to validity or the degree to which the measurement is a true and accurate reflection of the variable of interest.

b) **Balance :** Any measurement needs a framework or plan for its development. The extent to which the developed measure corresponds to the ideal measure reflects balance.

c) **Efficiency :** Basically we are looking for the greatest number of meaningful responses per unit of time. Gathering data costs time and money, so we can conserve our resources. A balance between time available to collect the data, cost, requirements for scoring and summarization and relevance should be sought.

d) **Objectivity :** Objectivity is a characteristic of the scoring or the assignment of meaning to the data rather than being a description of the method of data collection.

e) **Reliability :** Reliability is a complex characteristic but generally involves the idea of consistency of measurement. Consistency of measurement might be judged in terms of time, items, scores, examinees, examiners, or accuracy of classification.

f) **Fairness :** The criterion of fairness relates to a wide range of data characteristics ranging from freedom from bias to the administration o a test in a manner that allows everybody an equal chance to demonstrate their knowledge or skills.

Q10. What are the Functions of Measurement?

Ans :

1. It facilitates empirical descriptions of social and psychological phenomena.
2. Measurement renders data amenable to statistical manipulation and treatment.
3. Measurement facilitates testing of theories and hypotheses.
4. Measurement enables researchers to differentiate between objects or people interms of specific properties they process.

Q11. Outline the process of Measurement.

Ans :

1. **Identify the Concept of Interest :** Measurement begins by identifying a concept of interest for study. A concept is an abstract idea generalized from particular facts. A concept is a category of thought, or a category for grouping sense data together "As if they were all the same." All stoplight, regardless of location, become a broader concept. Thus, a concept is a category of thought.

2. **Develop a Construct :** Constructs are specific types of concepts that exist at higher

levels of abstraction. Constructs are invented for theoretical use. Generally, constructs are not directly observable. Instead, they are inferrer by some indirect method such as findings on a questionnaire. Examples of marketing constructs include brand loyalty, high-involvement purchasing, social class, personality, and channel power. Constructs aid researchers by simplifying and integrating the complex phenomena found in the marketing environment.

3. **Define the Concept Both Constitutively and Operationally :** The third and fourth steps in the research process are to first define the concept constitutively and the operationally. A constitutive (or theoretical or conceptual) definition defines a concept with other concepts and constructs, establishing boundaries for the construct under study; it states the central idea or concept under study. All constructs, to be scientifically useful, must be capable of being used in theories. A constitutively defined concept should fully distinguish the concept under investigation from all other concepts.

4. **Develop A Measurement Scale :** A scale is a set of symbols or numbers do constructed that the symbols or numbers can be assigned by a rule to the individual to whom the scale is applied. The assignment on the scale is indicated by the individual's possession of whatever the scale is supposed to measure. Creating a measurement scale begins with determining the level of measurement desirable or possible.

5. **Evaluate the Reliability and Validity of the Measurement :** An ideal market research study would provide information that is accurate, precise, lucid, and timely. Accurate data implies accurate measurement, or M = a, where :

   M refers to measurement A stands for complete accuracy.

   In market research, this ideal is rarely, if ever, obtained. Instead we have

$$M = A + E$$

Where

E = errors

Errors can be either random or systematic. Systematic error is error that results in a constant bias in the measurements. The bias results from faults in the measurement instrument or process. Random error also influences the measurements but not systematically. Thus, random error is transient in nature and does not occur in a consistent manner. A person may not answer a question truthfully because he is in a bad mood that day. Two scores on a measurement scale for a number of reasons.

Q12. Define scaling. Explain the properties of scales.

Ans :

Scales in marketing need not necessarily imply all the physical measures. For example, a rank scale does not possess additive property. It makes no sense to add two brands of ranks 1 and 2 to get a brand of rank 3.

Scaling is an advancement of the measuring concept that refers to the continuum on which, the objects to be measured are marked. While measurement is the actual assignment of numbers or symbols, scaling is the arrangement of these numbers in an ordeton a continuum. Such information helps the researcher in gaining the research objective of determining the stores image.

Properties of Scales

The measurement scales has the following properties,

(i) Distinctive Classification

Distinctives classification is the property of a measure which is used to differentiate objects or its characteristics into various categories. For example, gender categorizes the individuals into two different groups such as males and females.

(ii) Order

A measure that can be used to arrange the obj ects or their characteristics in a meaningful order is said to have order property.

Example

The arrangement of student's marks in an ascending or descending order.

(iii) Equal Distance

The measure is said to have equal distance property, if the difference between two consecutive categories of a measured scale are equal.

Example

The difference between the temperatures 40°C and 50°C is equal to the difference between temperatures, 60°C and 70°C i.e.,'1'.

(iv) Fixed Origin

A measurement scale used for measuring a characteristic is said to have the property of "Fixed origin' when there exists a 'meaningful zero' or absence of characteristic.

Example

Sales of a company where 'zero sales' describes that there is no sales or absence of sales.

Q13. What are the different types of measurement scales with examples?

Ans : **(Aug.-21, Nov.-20)**

a) Nominal Scale

Nominal scale is simply a system of assigning number symbols to events in order to label them. The usual example of this is the assignment of numbers of basketball players in order to identify them. Such numbers cannot be considered to be associated with an ordered scale for their order is of no consequence; the numbers are just convenient labels for the particular class of events and as such have no quantitative value. Nominal scales provide convenient ways of keeping track of people, objects and events. One cannot do much with the numbers involved. For example, one cannot usefully average the numbers on the back of a group of football players and come up with a meaningful value. Neither can one usefully compare the numbers assigned to one group with the numbers assigned to another. The counting of members in each group is the only possible arithmetic operation when a nominal scale is employed. Accordingly, we are restricted to use mode as the measure of central tendency. There is no generally used measure of dispersion for nominal scales. Chi-square test is the most common test of statistical significance that can be utilized, and for the measures of correlation, the contingency coefficient can be worked out.

Nominal scale is the least powerful level of measurement. It indicates no order or distance relationship and has no arithmetic origin. A nominal scale simply describes differences between things by assigning them to categories. Nominal data are, thus, counted data. The scale wastes any information that we may have about varying degrees of attitude, skills, understandings, etc. In spite of all this, nominal scales are still very useful and are widely used in surveys and other ex-post-facto research when data are being classified by major sub-groups of the population.

(b) Ordinal Scale

The lowest level of the ordered scale that is commonly used is the ordinal scale. The ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule. Rank orders represent ordinal scales and are frequently used in research relating to qualitative phenomena. A student's rank in his graduation class involves the use of an ordinal scale. One has to be very careful in making statement about scores based on ordinal scales. For instance, if Ram's position in his class is 10 and Mohan's position is 40, it cannot be said that Ram's position is four times as good as that of Mohan. The statement would make no sense at all. Ordinal scales only permit the ranking of items from highest to lowest. Ordinal measures have no absolute values, and the real differences between adjacent ranks may not be equal. All that can be said is that one person is higher or lower on the scale than another, but more precise comparisons cannot be made.

Thus, the use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without our being able to state how much greater or less. The real difference between ranks 1 and 2 may be more or less than the difference between ranks 5 and 6. Since the numbers of this scale have only a rank meaning, the appropriate measure of central tendency is the median. A percentile or quartile measure is used for measuring dispersion. Correlations are restricted to various rank order methods. Measures of statistical significance are restricted to the non-parametric methods.

(c) Interval Scale

In the case of interval scale, the intervals are adjusted in terms of some rule that has been established as a basis for making the units equal. The units are equal only in so far as one accepts the assumptions on which the rule is based. Interval scales can have an arbitrary zero, but it is not possible to determine for them what may be called an absolute zero or the unique origin. The primary limitation of the interval scale is the lack of a true zero; it does not have the capacity to measure the complete absence of a trait or characteristic. The Fahrenheit scale is an example of an interval scale and shows similarities in what one can and cannot do with it. One can say that an increase in temperature from 30° to 40° involves the same increase in temperature as an increase from 60° to 70°, but one cannot say that the temperature of 60° is twice as warm as the temperature of 30° because both numbers are dependent on the fact that the zero on the scale is set arbitrarily at the temperature of the freezing point of water. The ratio of the two temperatures, 30° and 60°, means nothing because zero is an arbitrary point.

Interval scales provide more powerful measurement than ordinal scales for interval scale also incorporates the concept of equality of interval. As such more powerful statistical measures can be used with interval scales. Mean is the appropriate measure of central tendency, while standard deviation is the most widely used measure of dispersion. Product moment correlation techniques are appropriate and the generally used tests for statistical significance are the 't' test and 'F' test.

(d) Ratio scale

Ratio scales have an absolute or true zero of measurement. The term 'absolute zero' is not as precise as it was once believed to be. We can conceive of an absolute zero of length and similarly we can conceive of an absolute zero of time. For example, the zero point on a centimeter scale indicates the complete absence of length or height. But an absolute zero of temperature is theoretically unobtainable and it remains a concept existing only in the scientist's mind. The number of minor traffic-rule violations and the number of incorrect letters in a page of type script represent scores on ratio scales. Both these scales have absolute zeros and as such all minor traffic violations and all typing errors can be assumed to be equal in significance. With ratio scales involved one can make statements like "Jyoti's" typing performance was twice as good as that of "Reetu." The ratio involved does have significance and facilitates a kind of comparison which is not possible in case of an interval scale.

Ratio scale represents the actual amounts of variables. Measures of physical dimensions such as weight, height, distance, etc. are examples. Generally, all statistical techniques are usable with ratio scales and all manipulations that one can carry out with real numbers can also be carried out with ratio scale values. Multiplication and division can be used with this scale but not with other scales mentioned above. Geometric and harmonic meaqs can be used as measures of central tendency and coefficients of variation may also be calculated.

1.5.2 Reliability and Validity in Measurement of Variables

Q14. Explain the test of reliability of measurement.

Ans :

The test of reliability is the important test of sound measurement. A measuring instrument is reliable if it provides consistent results. Reliable measuring instrument does contribute to validity, but a reliable instrument need not be a valid instrument. For instance, a scale that consistently overweights objects by five kgs., is a reliable scale, but it does not give a valid measure of weight. But the otherway is not true i.e., a valid instrument is always reliable . Accordingly reliability is not as valuable as validity, but it is easier to assess reliability in comparison to validity. If the quality of reliability is satisfied by an instrument, then while using it we can be confident that the transient and situational factors are not interfering.

Two aspects of reliability viz., stability and equivalence deserve special mention. The stability aspect is concerned with securing consistent results with repeated measurements of the same person and with the same instrument. We usually determine the degree of stability by comparing the results of repeated measurements. The equivalence aspect considers how much error may get introduced by different investigators or different samples of the items being studied. A good way to test for the equivalence of measurements by two investigators is to compare their observations of the same events. Reliability can be improved in the following two ways:

(i) By standardising the conditions under which the measurement takes place i.e., we must ensure that external sources of variation such as boredom, fatigue, etc., are minimised to the extent possible. That will improve stability aspect.

(ii) By carefully designed directions for measurement with no variation from group to group, by using trained and motivated persons to conduct the research and also by broadening the sample of items used. This will improve equivalence aspect.

Q15. Explain the test of validity of measurement.

(OR)

What are the different types of test of validity of measurement?

Ans :

Validity is the most critical criterion and indicates the degree to which an instrument measures what it is supposed to measure. Validity can also be thought of as utility. In other words, validity is the extent to which differences found with a measuring instrument reflect true differences among those being tested. But the question arises: how can one determine validity without direct confirming knowledge? The answer may be that we seek other relevant evidence that confirms the answers we have found with our measuring tool. What is relevant, evidence often depends upon the nature of the research problem and the judgement of the researcher. But one can certainly consider three types of validity in this connection:

(i) Content validity;

(ii) Criterion-related validity and

(iii) Construct validity.

(i) Content Validity

Content validity is the extent to which a measuring instrument provides adequate coverage of the topic under study. If the instrument contains a representative sample of the universe, the content validity is good. Its determination is primarily judgemental and intuitive. It can also be determined by using a panel of persons who shall judge how well the measuring instrument meets the standards, but there is no numerical way to express it.

(ii) Criterion-related Validity

Criterion-related validity relates to our ability to predict some outcome or estimate the existence of some current condition. This form of validity reflects the success of measures used for some empirical estimating purpose. The concerned criterion must possess the following qualities:

(a) **Relevance:** A criterion is relevant if it is defined in terms we judge to be the proper measure.

(b) **Freedom from bias:** Freedom from bias is attained when the criterion gives each subject an equal opportunity to score well.

(c) **Reliability:** A reliable criterion is stable or reproducible.

(d) **Availability:** The information specified by the criterion must be available.

In fact, a Criterion-related validity is a broad term that actually refers to,

(i) Predictive validity and

(ii) Concurrent validity

The former refers to the usefulness of a test in predicting some future performance whereas the latter refers to the usefulness of a test in closely relating to other measures of known validity. Criterion-related validity is expressed as the coefficient of correlation between test scores and some measure of future performance or between test scores and scores on another measure of known validity.

(iii) Construct Validity

Construct validity is the most complex and abstract. A measure is said to possess construct validity to the degree that it confirms to predicted correlations with other theoretical propositions. Construct validity is the degree to which scores on a test can be accounted for by the explanatory constructs of a sound theory. For determining construct validity, we associate a set of other propositions with the results received from using our measurement instrument. If measurements on our devised scale correlate in a predicted way with these other propositions, we can conclude that there is some construct validity.

If the above stated criteria and tests are met with, we may state that our measuring instrument is valid and will result in correct measurement; otherwise we shall have to look for more information and/or resort to exercise of judgement.

1.5.3 Sources of Error in Measurement

Q16. What are the various sources of errors in measurement.

Ans : **(Imp.)**

Sources of Error in Measurement

Measurement should be precise and unambiguous in an ideal research study. This objective, however, is often not met with in entirety. As such the researcher must be aware about the sources of error in measurement. The following are the possible sources of error in measurement.

(a) Respondent

At times the respondent may be reluctant to express strong negative feelings or it is just possible that he may have very little knowledge but may not admit his ignorance. All this reluctance is likely to result in an interview of 'guesses.' Transient factors like fatigue, boredom, anxiety, etc. may limit the ability of the respondent to respond accurately and fully.

(b) Situation

Situational factors may also come in the way of correct measurement. Any condition which places a strain on interview can have serious effects on the interviewer-respondent rapport. For instance, if someone else is present, he can distort responses by joining in or merely by being present. If the respondent feels that anonymity is not assured, he may be reluctant to express certain feelings.

(c) Measurer

The interviewer can distort responses by rewording or reordering questions. His behaviour, style and looks may encourage or discourage certain replies from respondents. Careless mechanical processing may distort the findings. Errors may also creep in because of incorrect coding, faulty tabulation and/or statistical calculations, particularly in the data-analysis stage.

(d) Instrument

Error may arise because of the defective measuring instalment. The use of complex words, beyond the comprehension of the respondent, ambiguous meanings, poor pnnthig, inadequate space for replies, response choice omissions, etc. are a few things that make the measuring instrument defective and may result in measurement errors. Another type of instrument deficiency is the poor sampling of the universe of items of concern.

Researcher must know that correct measurement depends on successfully meeting all of the problems listed above. He must, to the extent possible, try to eliminate, neutralize or otherwise deal with all the possible sources of error so that the final results may not be contaminated.

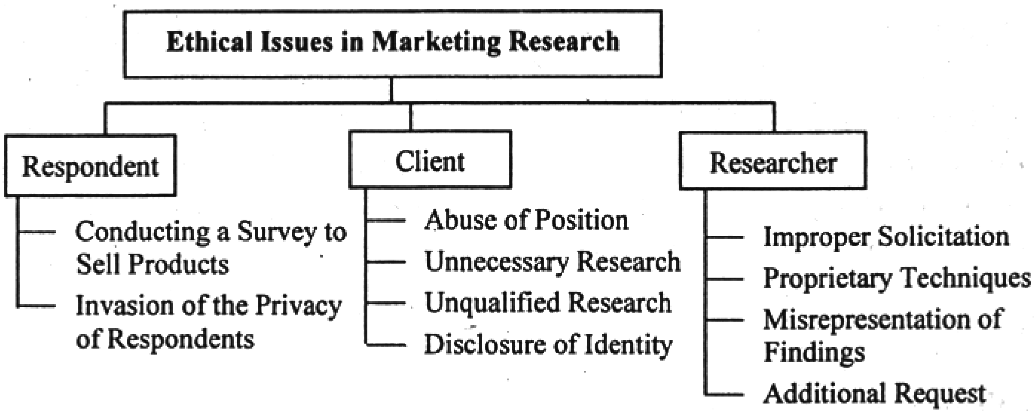
1.6 Ethics in Business Research

Q17. What do you understand by ethics in business research?

Ans : **(Imp.)**

Ethics are moral principles or values generally governing the conduct of an individual or group. Ethical questions range from practical, narrowly defined issues, such as a researcher's obligation to be honest with its customers, to broader social and philosophical questions, such as a company's responsibility to preserve the environment and protect employee rights.

Unethical practices by some suppliers include abusing respondents, selling unnecessary research, and violating client include requesting bids when a supplier has been predetermined, requesting bids gain to free advise methodology, marketing false promises, and issuing unauthorised requests for proposals. Marketing research field services have used professional respondents, which is unethical.



In market research there are following stakeholders, and the ethical issues relating to these stakeholders are as follows:

1. Respondent

These are one of the major stakeholders because without their involvement the research is not possible. Therefore, market researchers should protect the respondents from unethical research practices. The following will show the respondent-researcher relationship.

i) **Conducting a Survey to Sell Products:** Respondents have been deliberately deceived by researchers. Some unethical marketers have been known to tell respondents that they were conducting a survey which actually lead-in to a sales presentation, or to get information that could be used for sales leads or mailing lists. This is called sugging in trade language. It is illegal as well as unethical.

   These practices violate the respondents trust and erode their willingness to support surveys.

ii) **Invasion of the Privacy of Respondents:** The privacy of the respondent is a legal right and hence it must be protected. It means that respondents, who are promised with the anonymity or confidentiality, should be given the same. Respondents should also not be contacted at times which are inconvenient for them. Another concern is also the buying and selling of mailing lists through deceptive means also forms a part of these unethical practices.

2. Client

The complexity that is surrounding the research-client relationship deserves special attention from an ethical standpoint.

i) **Abuse of Position:** Since the researcher possesses the research expertise, the researcher has a responsibility not to take unfair advantage of his position. The researcher should make every effort to follow correct research procedures, adopt a suitable approach and research design. In short, the researcher must conduct quality research while respecting the clients' resources of time and money.

ii) **Unnecessary Research:** The researcher has the ethical duty not to perform unnecessary research. For example, primary research is not required if secondary data provide the necessary information.

iii) **Unqualified Research:** The researcher sometimes may not possess technical expertise that is required for the research. Thus, it is his duty to let the client know these limitations and refiise the project.

iv) **Disclosure of Identity:** The client has the right to expect that its identity will be protected before, during, and after the completion of the project. The researcher is ethically bound not to reveal the client's identity to competitors, respondents, etc., without the consent of the client involved.

3. Researcher: The researcher or the research firm has the right to be treated ethically as well. Ethical treatment by clients involves several issues.

i) **Improper Solicitation:** When a research firm submits a proposal to a prospective client, it should be confident that the client is seriously considering employing it to conduct the research projects. The client should not make an attempt to misuse the first proposal by turning it over to another firm for execution. The research proposal is the property or the research firm.

ii) **Proprietary Techniques:** The researcher has the right to expect that the proprietary techniques will not be revealed by the client to other research firms. The client should also refrain from using these techniques in future without prior permission by the researcher who has developed the technique.

iii) **Misrepresentation of Findings:** The client should not distort the research findings to their own benefit at the expense of the researcher's reputation. The client should reveal the truth and nothing else.

iv) **Additional Request:** The client assigning the project to a particular researcher may request him to provide some additional information at the original project cost. This extra information involves additional expenditure of the researcher.

Q18. Discuss the various ethical issues in business research.

Ans : **(Nov.-21)**

The ethical issues can be observed at various levels. The different components or levels of a research process with their ethical issues are discussed below,

1. Sponsor/solicitor
2. Consultant/researcher
3. Researcher team members
4. Participants.

1. Sponsor/Solicitor

The faculty member who is accountable for assigning the topic for the study is known as the 'sponsor'. He can decide the subject as per his perception of the subject, decisions made by the management of the institute etc.,

Advantages

(i) The purpose must be clear without any hidden objective.

(ii) The sponsor must be clear during the settlement of terms and condition.

(iii) The sponsor must not accept the report as per his convenience.

Disadvantages

(i) One should avoid conducting a research study which may help in bringing favourable points for increasing the stock price of a firm.

(ii) It is unethical to design and conduct a survey to contrive the results which are already justified by the decision maker.

(iii) It is a wrong practice to manipulate the data by truncating the data in order to influence the results in a favourable direction.

2. Consultant/Researcher

The group of students who work as a team on an assigned topic are known as researchers. The ethical issues related to the consultant or researcher are as follows,

(i) The research should be carried out in a professional way by maintaining the privacy and confidentiality of the research data and outcomes.

(ii) The research process should be organized at all levels by consulting the management and the employees.

(iii) Making sure that the sponsoring organization provides fair and true state of affairs.

(iv) Suitable tools should be used for the scope and objective of the research.

(v) Full freedom must be given to the employees for providing their views without any restriction.

3. Researcher Team Members

Research team members are the individuals who actually conduct the study. The ethical issues related to the researcher are as follows,

(i) Assignment of responsibilities as per the need of expertise.

(ii) Internal data to be collected and efforts must be taken to achieved required data for the firm.

(iii) Suitable and reliable data must be collected from external sources by considering their importance, limitations and scope.

4. Participants/Subjects/Respondents

When the research requires collection of primary data from other entities (i.e., retail outlets) and individuals (i.e., customers) by communicating with them, these entities and individuals are referred to as respondents/ participants/subjects.

Q19. What are the various ethical issues at various levels of research process ?

Ans :

There exist various ethical issues at various levels of research process. Each step of research process along with its related ethical issues are discussed as follows,

1. Objective

The objective should be clear without any hidden plan. It should induce interest and enthusiasm in researchers and participants.

2. Defining Problem

The problem should be described in such a way that it flows directly from the objectives and should be like a challenge for both researchers and respondents.

3. Research Design

The researcher should avoid selecting improper design for simplifying the research study.

4. Sampling

The sample must be the model of population from which the outcomes are to be drawn. Likewise, the sample size must be suitable enough to extract the valid outcomes. As sampling often results in unethical research, one should think before believing in any outcomes.

5. Questionnaire Design

It is important to ensure that the scale technique used to record the opinions of participants should be ethical. It should not involve any negative or positive description. The researcher should not produce a document that collects the data from various parties and serve it to more than one client without informing the other concerned parties. It is unethical to avail profits using this method.

6. Data Collection

The data must be collected sincerely without any unethical practice. It is essential to avoid unethical practices such as bias in data collection, collecting only the data which is convient to record, filling questionnaire without interviewing.

7. IT Ethics

Information technology involve many ethical issues due to insincerity and unreliability in the IT dept personnel. Honesty and trustworthy personnel should be employed to ensure reliability.

8. Data Presentation

The data should be presented in such a way that it should not involve any visual deception.

9. Data Analysis

Analysis should be carried out to check if the data is as per the planned methodology or not. Any unethical errors involved should be corrected.

10. Interpreting Conclusions and Report Writing

Interpretations must be drawn based on the conclusions made in a study with valid arguments.

Short Question and Answers

1. Research

Ans :

Research is a scientific inquiry aimed at learning new facts, testing ideas, etc. It is the systematic collection, analysis and interpretation of data to generate new knowledge and answer a certain question or solve a problem.

Research is the systematic process of collecting and analyzing information to increase our understanding of the phenomenon under study. It is the function of the research to contribute to the understanding of the phenomenon and to communicate that understanding to others.

Definitions of Research

According to Clifford Woody "Research comprises of defining and redefining problems, formulating hypothesis or suggested solutions, making deductions and reaching conclusions; and a last carefully testing the conclusions to determine whether they fit the formulating hypothesis".

According to Redman and Mory "Research is a systematized effort to gain new knowledge".

According to Advanced Learner's Dictionary "Research is a careful investigation or inquiry especially through search for new facts in any branch of knowledge.

2. Characteristics of Research

Ans :

i) **Empirical:** Research is based on direct experience or observation by the researcher.

ii) **Logical:** Research is based on valid procedures and principles.

iii) **Cyclical:** Research is a cyclical process because it starts with a problem and ends with a problem.

iv) **Analytical:** Research utilizes proven analytical procedures in gathering the data, whether historical, descriptive, experimental and case study.

v) **Critical:** Research exhibits careful and precise judgment.

vi) **Methodical:** Research is conducted in a methodical manner without bias using systematic method and procedures.

3. Define the term variable.

Ans :

A variable can be simply defined as any factor that varies. It is a certain aspect of an experimental condition that can vary or take on different characteristics in different conditions. The values of variables can change at different times for the same person or object or at the same time for different persons or objects. They are vital because they help to focus on specific events out of the many that are associated with the phenomenon. They connect the empirical world with the theoretical world. Health status, age, attitudes, memory, production units, motivation etc., are some of the examples of variables.

4. Compare and contrast between independent and dependent variable.

Ans :

The differences between independent and dependent variables are follows,

Independent variable	Dependent variable
1. It is responsible for bringing about change(s) in a phenomenon.	1. It is the outcome of the change (s) brought about by the introduction of the independent variable.
2. It is controlled and manipulated by the researcher.	2. It is a measurable variable which is difficult to manipulate.
3. It takes the form of experiment stimulus having two attributes i.e., present or absent.	3. It has attributes which are indirect, direct or through constructs.
4. It is a "casual" variable.	4. It is considered as a "caused" variable
5. It appears, disappears, or varies as the researcher introduces, eliminates or varies it.	5. It is observed and measured to determine the effect of independent variable.
6. It is the presumed cause, changes in which leads to changes in the status of dependent variable.	6. It is the presumed effect (response) in which the researcher is interested.

5. Define measurement.

Ans :

Measurement is the process of assigning numbers or labels to objects, persons, states, or events in accordance with specific rules to represent quantities or qualities of attributes. Measurement then, is a procedure used to assign numbers that reflect the amount of an attribute possessed by an event, person, or object. Note that the event, person, or object is not being measured, but rather its attributes are. A researcher, for example, does not measure a consumer but measures attitude, income, brand loyalty, age, and other relevant factors.

Definition of Measurement

According to stevens measurement is the assignment of numerals to objects or events according to rules.

"**According to Campbell** measurement is the assignment of numbers to represent properties."

6. What are the Functions of Measurement?

Ans :

1. It facilitates empirical descriptions of social and psychological phenomena.
2. Measurement renders data amenable to statistical manipulation and treatment.
3. Measurement facilitates testing of theories and hypotheses.
4. Measurement enables researchers to differentiate between objects or people interms of specific properties they process.

7. Define scaling.

Ans :

Scales in marketing need not necessarily imply all the physical measures. For example, a rank scale does not possess additive property. It makes no sense to add two brands of ranks 1 and 2 to get a brand of rank 3.

Scaling is an advancement of the measuring concept that refers to the continuum on which, the objects to be measured are marked. While measurement is the actual assignment of numbers or symbols, scaling is the arrangement of these numbers in an ordeton a continuum. Such information helps the researcher in gaining the research objective of determining the stores image.

8. Ordinal Scale

Ans :

The lowest level of the ordered scale that is commonly used is the ordinal scale. The ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule. Rank orders represent ordinal scales and are frequently used in research relating to qualitative phenomena. A student's rank in his graduation class involves the use of an ordinal scale. One has to be very careful in making statement about scores based on ordinal scales. For instance, if Ram's position in his class is 10 and Mohan's position is 40, it cannot be said that Ram's position is four times as good as that of Mohan. The statement would make no sense at all. Ordinal scales only permit the ranking of items from highest to lowest. Ordinal measures have no absolute values, and the real differences between adjacent ranks may not be equal. All that can be said is that one person is higher or lower on the scale than another, but more precise comparisons cannot be made.

Thus, the use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without our being able to state how much greater or less. The real difference between ranks 1 and 2 may be more or less than the difference between ranks 5 and 6. Since the numbers of this scale have only a rank meaning, the appropriate measure of central tendency is the median. A percentile or quartile measure is used for measuring dispersion. Correlations are restricted to various rank order methods. Measures of statistical significance are restricted to the non-parametric methods.

9. Validity

Ans :

Validity is the most critical criterion and indicates the degree to which an instrument measures what it is supposed to measure. Validity can also be thought of as utility. In other words, validity is the extent to which differences found with a measuring instrument reflect true differences among those being tested. But the question arises: how can one determine validity without direct confirming knowledge? The answer may be that we seek other relevant evidence that confirms the answers we have found with our measuring tool. What is relevant, evidence often depends upon the nature of the research problem and the judgement of the researcher. But one can certainly consider three types of validity in this connection:

(i) Content validity;

(ii) Criterion-related validity and

(iii) Construct validity.

10. Ethics in business research

Ans :

Ethics are moral principles or values generally governing the conduct of an individual or group. Ethical questions range from practical, narrowly defined issues, such as a researcher's obligation to be honest with its customers, to broader social and philosophical questions, such as a company's responsibility to preserve the environment and protect employee rights.

Unethical practices by some suppliers include abusing respondents, selling unnecessary research, and violating client include requesting bids when a supplier has been predetermined, requesting bids gain to free advise methodology, marketing false promises, and issuing unauthorised requests for proposals. Marketing research field services have used professional respondents, which is unethical.

UNIT II	**Research design -** Research Problem- purpose of Research design, Types of Research Design- Experimental research design, Research Design for cross sectional, longitudinal studies, Research design for action research – Characteristics of the good research design. **Data Collection Methods & Tools:** Types of Data, Sources and Instruments for data, Guidelines for questionnaire, Sampling and its application.

2.1 Research Problem

Q1. What is Research Problem? Explain the various conditions involved in research problem.

Ans :

A research problem, in general, refers to some difficulty which a researcher experiences in the context of either a theoretical or practical situation and wants to obtain a solution for the same. Usually we say that a research problem does exist if the following conditions are met with:

(i) There must be an individual (or a group or an organization), let us call it I,' to whom the problem can be attributed. The individual or the organization, as the case may be, occupies an environment, say N', which is defined by values of the uncontrolled variables, Y_j.

(ii) There must be at least two courses of action, say C_l and C_2, to be pursued. A course of action is defined by one or more values of the controlled variables. For example, the number of items purchased at a specified time is said to be one course of action.

(iii) There must be at least two possible outcomes, say O_1 and O_2 of the course of action, of which one should be preferable to the other. In other words, this means that there must be at least one outcome that the researcher wants, i.e., an objective.

(iv) The courses of action available must provides some chance of obtaining the objective, but they cannot provide the same chance, otherwise the choice would not matter. Thus, if $P(O_j \backslash I, C_j, N)$ represents the probability that an outcome O. will occur, if I select C. in N, then $P(O_1 \backslash I, C_1, N) \neq P(O_1 \backslash I, C_2, N)$. In simple words, we can say that the choices must have unequal efficiencies for the desired outcomes.

Over and above these conditions, the individual or the organisation can be said to have the problem only if I does not know what course of action is best, i.e., I, must be in doubt about the solution. Thus, an individual or a group of persons can be said to have a problem which can be technically described as a research problem, if they (individual or the group), having one or more desired outcomes, are confronted with two or more courses of action that have some but not equal efficiency for the desired objective(s) and are in doubt about which course of action is best.

We can, thus, state the components' of a research problem as under:

(i) There must be an individual or a group which has some difficulty or the problem.

(ii) There must be some objective(s) to be attained at. If one wants nothing, one cannot have a problem.

(iii) There must be alternative means (or the courses of action) for obtaining the objective(s) one wishes to attain. This means that there must be *at least two means* available to a researcher for if he has no choice of means, he cannot have a problem.

(iv) There must remain some doubt in the mind of a researcher with regard to the selection of alternatives. This means that research must answer the question concerning the relative efficiency of the possible alternatives.

(v) There must be some environment(s) to which the difficulty pertains.

Thus, a research problem is one which requires a researcher to find out the best solution for the given problem, i.e., to find out by which course of action the objective can be attained optimally in the context of a given environment. There are several factors which may result in making the problem complicated.

For instance, the environment may change affecting the efficiencies of the courses of action or the values of the outcomes.

The number of alternative courses of action may be very large; persons not involved in making the decision may be affected by it and react to it favourably or unfavourably, and similar other factors. All such elements (or at least the important ones) may be thought of in context of a research problem.

Q2. What is the necessity of defining a research problem? Explain.

Ans :

The statement signifies the need for defining a research problem. The problem to be investigated must be defined unambiguously for that will help to discriminate relevant data from the irrelevant ones.

A proper definition of research problem will enable the researcher to be on the track whereas an ill-defined problem may create hurdles. Questions like: What data are to be collected?

What characteristics of data are relevant and need to be studied? What relations are to be explored.

What techniques are to be used for the purpose? and similar other questions crop up in the mind of the researcher who can well plan his strategy and find answers to all such questions only when the research problem has been well defined.

Thus, defining a research problem properly is a prerequisite for any study and is a step of the highest importance. In fact, formulation of a problem is often more essential than its solution.

It is only on careful detailing the research problem that we can work out the research design and can smoothly carry on all the consequential steps involved while doing research.

2.2 Research Design

Q3. Define Research Design. What are the components of Research Design?

Ans : **(Jan.-20, Imp.)**

A research design is a framework or blueprint for conducting the marketing research project. It details the procedures necessary for obtaining the information needed to structure or solve marketing research problems. In simple words it is the general plan of how you will go about your research.

Definitions of Research Design

According to Kerlinger, Research design is the plan, structure and strategy of investigation conceived so as to obtain answers to research questions and to control variance.

According to Green and Tull, A research is the specification of methods and procedures for acquiring the information needed. It is the overall operational pattern or framework of the project that stipulates what information is to be collected from which sources by what procedures.

Components of Research Design

1. **Title of the Study:** Should be brief, precise, and project the scope in generalized terms.
2. **Statement of the Problem:** Should be unambiguous, precise, and the usage should be clear, simple, and concise.
3. **Review of Previous Studies:** Should be a brief survey of a relevant literature in the subject concerned and presented subject wise and reviewed critically and pinpoint the stage from where further research is called for.
4. **Definition of Concept or Theoretical Principles involved if any:** Should be defined in general terms and be linked with the study.
5. **Coverage and Scope of the Study:** Should consider, geographical, temporal, and functional dimensions.
6. **Objectives of the Study:** Should explain the main purpose precisely and may be in the form of questions or an explanation to a particular issue.

7. **Formulation of Hypothesis:** Should be empirical, conceptually clear, specific, close to things observable and related to the body of theory.

8. **Methods of Investigation:** It depends upon the nature of study but researcher should define survey methodology, and statistical techniques adopted.

9. **Sampling Design:** Definition of the universe or population, size of the sample, and representativeness of the sample should be defined.

10. **Constructing of Schedule or Questionnaire :** Questions should be in an order on a form. Open ended questions are designed to permit a free response. The questions to be asked should have a direct bearing on the problem, avoiding the personal questions and multiple meanings.

11. **Data Collection:** Depends up the subject matter, the unit of enquiry and the scope of the study.

12. **Analysis of Data:** To fulfill the objective or hypothesis the researcher should analyze the data subject to the appropriate statistical analysis besides tabulation. Tabulation of results in a meaningful way is by itself a technique and an art. The data given in the tables must be in a self-explanatory form.

13. **Interpretation of Results:** Researcher should draw inferences based on usual test for significance and relate with previous findings, to a wider field of generalizations, to scientific objectivity, and to uncover any additional factors which would not be visualized by the investigator earlier.

14. **Reporting the Findings:** Should be clear, specific, simple, and directly relating to the objective of the study. Researcher must report that what has been discovered or innovated to fulfill the need for which the study was taken up and to ensure proper directions to other researchers in carrying out similar researches.

Q4. Explain the steps involved in Research Design.

Ans :

Steps in Research Design Process

1. **Determining sources of data**

The first step is to determine the sources of data to be used. The marketing researcher has to decide whether he has to collect data or depends on exclusively on secondary data. Some times, the research study is depends on both secondary and primary. Some times he can depends upon collecting fresh date.

2. **Designing data collection forms**

Once the decision in favour of collection of primary data is taken, one has to decide the mode of collection. There are two methods available are

a) Observational method

b) Survey method

a) Observation method : This method suggests that data are collected through one's observation. If the researcher is keen observer, with integrity he would be in a position to observe and record data faithfully and accurately. While the observation method may be suitable incase of some studies, several things of interest such as attitudes, opinions, motivations and other intangible states of mind can not observed.

b) Survey method : In marketing research, field surveys are commonly used to primary data form the respondents. This survey can be : 1) personal 2) telephonic 3) by mail 4) by diary

Personal and mail surveys are more frequently used in India. Telephonic survey is suitable when very limited information is sought in a short period of time personal interviews are suitable when detailed information is to be collected. Sometimes a combination of two or more methods could also be used.

3. **Determining sample design and sample size**

   Another aspect which forms a part of research process is the sampling plant. When the marketing researcher has decided to carry out a field survey has to decide where it is to be a simple survey. In almost all cases, a sample survey is undertaken on account of overwhelming advantages over a census survey.

4. **Organizing and conducting the field survey**

   The next step is to organize and conduct the field survey. Two important aspects should be looked into interviewing and the supervision of field work.

   Interview task is very simple but it very difficult in marketing research. Supervision of field work is equal important to ensure timely and proper completions of the field survey.

5. **Processing and analyzing the collected data**

   Once the field survey is over and question have been received the next task is to aggregate the data in a meaning full manner. A number of tables are prepared to bring out the main characteristics of the data.

   The researcher should have a well thought out frame work for processing and analyzing data, and this should be done prior to the collection.

   While designing the research study, the researcher should give adequate thought to the use of particular analytical techniques.

6. **Preparing the research report**

   Once the data has been tabulated. Interpreted and analyzed, the marketing researcher is require and to prepare his report embodying the findings of the research study and recommendations.

   The research should follow the main principles of writing a report, some of these principles are objectivity, clarity in the presentation of ideas and use of chart and diagrams. The essence of a good research report is that it effectively communicates its research findings.

   A research design may be its successful implementation depend in no small measure on its managers. In fact the management research most important marketing.

Q5. What are the Requirements for a Good Research Design.

Ans :

1. **Well Defined Problems:** Nature and scope of the problem to be studied must be stated clearly, or say must be well defined and formulated.
2. **Clarity in Formulation:** If any hypothesis is to be tested it must be clearly formulated.
3. **Testable:** The research design must adequately answer the research questions. And test the hypothesis.
4. **Identifying Variables:** Relevant variables must be clearly identified and operationalized. Adequate methods of collecting the information and methods of logically deriving the conclusion must be developed. Only then control of variance is possible.
5. **Serve the Validity Needs:** The research design must be structured in a manner that it fulfils the need of internal and external validity.

2.2.1 Purpose of Research Design

Q6. Explain the Purpose of Research Design.

Ans :

A research design is simply the framework or plan for a study that is used as a guide in collecting and analyzing the data. It is a blueprint that is followed in completing a study. Research design is the blue print for collection measurement and analysis of data. Actually it is a map that is usually developed to guide the research.

Purpose of a Research Design

Research designs are used for the following purposes;

(i) To minimize the Expenditure

Research design carries an important influence on the reliability of the results attained .It therefore provides a solid base for the whole research. This makes the research as effective as possible by providing maximum information with minimum spending of effort, money and time by preparing the advance plan of all about the research.

(ii) To Facilitate the Smooth Scaling

Research design is needed because it facilitates the smooth scaling of the various research operations, thereby making research as efficient as possible yielding maximal information with minimal expenditure of effort, time and money.

(iii) To Collect the Relevant Data and Technique

Research design stands for advance planning of the methods to be adopted for collecting the relevant data and the techniques to be used in their analysis, keeping in view the objective of the research and the availability of staff time and money. Poor preparation of research design upset the entire project.

(iv) To Provide Blue Print for Plans

Research design is needed due to the fact that it allows for the smooth working of many research operations. It is like blue print which we need in advance to plan the methods to be adopted for collecting the relevant data and techniques to be used in its analysis for preparation of research project. Just as for better economical and attractive construction of a house need a blue print and a map of that, similarly we needs a blue print or a design for the smooth flow of operation of research.

(v) To Provide an Overview to other Experts

A research design provides an overview of all the research process and with the help of the design we can take the help and views of experts of that field .The design helps the investigator to organize his ideas , which helps to recognize and fix his faults.

(vi) To provide a direction

A research design provides a proper or particular direction to the other executives and others who are helping us into the process. The researcher studies available, literature and learns about new alternative approaches.

2.3 Types of Research Design Research Design for Cross Sectional, Longitudinal Studies

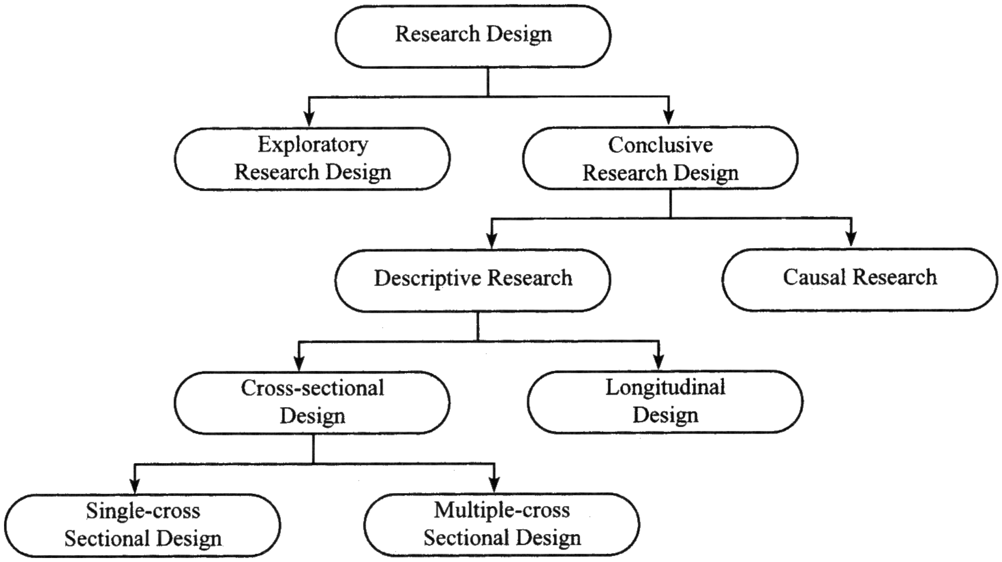
Q7. Discuss in detail classification/types of research designs.

(or)

What are the different types of research designs.

Ans : **(Nov.-20, Jan.-20, Imp.)**

The classification of research designs is explained in the following figure,



The research design is classified into two types,

I) Exploratory research design

II) Conclusive research design.

I) Exploratory Research Design

Exploratory research is concerned with discovering the general nature of the problem and the variables that are related to it. Exploratory research is characterized by a high degree of flexibility, and it tends to rely on secondary data, judgement samples, small-scale surveys or simple experiments, case analyses, and subjective evaluation of the results.

Methods of Exploratory Research

There are several methods of exploratory research. They are,

1. Experience surveys
2. Secondary data analysis
3. Case studies
4. Pilot studies.

1. Experience Surveys

It is an exploratory research technique in which experienced and knowledgeable individuals such as wholesalers and retailer's, opinions are considered about the existing research problem. The decision maker personally takes sometime to analyze the situation. This is also called as "informal experience survey". This method is used only to formulate the problem but it is nothing to do with decision making.

Merits of Experience Surveys

Experience surveys has the following merits,

(i) It helps to formulate the problem

(ii) It helps to clarify the concepts

(iii) It helps to get ideas about the problem

(iv) It is used to collect and interpret the available information which is found to be less expensive than other primary data.

Demerits of Experience Surveys

It comprises of the following demerits,

(i) It provides mostly the qualitative information.

(ii) As it includes interviews, opinions of different individuals which may create a confusion in the minds of the marketing managers. It restricts them in taking quick decisions.

2. Secondary Data Analysis

Drawing information from secondary sources or data (e.g: literatures) also constitutes to exploratory research. The secondary data is quite important in the process of applied research. The data that has been compiled is one of the most frequent forms of exploratory research.

Merits of Secondary Data Analysis

Some of the merits of secondary data analysis are,

(i) The secondary data is important to marketing managers who conduct the situation analysis.

(ii) The secondary data is frequently used in applied research process.

(iii) Through this technique, several conclusions (or decisions) can be drawn.

Demerits of Secondary Data Analysis

The following are some of the demerits of secondary data analysis,

(i) The findings from this technique can be vague.

(ii) Most exploratory techniques make use of smaller samples, which may not be sufficient to describe the characteristics of entire population.

3. Case Studies

The researchers conduct the case study method to obtain information from such situations which are quite similar to the problem situation of a research.

Merits of Case Studies

Case studies surveys the following merits,

(i) The primary advantage of case study is that the whole organization can be subjected to in-depth analysis.

(ii) The case study helps in focussing on the problems and provides solution.

Demerits of Case Studies

It comprises of the following demerits,

(i) The results from the exploratory research should be taken as uncertain.

(ii) Acquiring the information about competitors is quite difficult because everyone likes to keep the secret of their success with themselves.

4. Pilot Studies

Pilot study includes various techniques of research. Sometimes, researchers of pilot studies examine the consumer experience to gain valuable insight about the customers and their behaviour without revealing their positions with the company. Such methods are useful for revealing those situations that requires investigations.

Merits of Pilot Studies

Pilot studies has the following merits,

(i) The major benefit is that it generates an insight into customer's experiences and clarifies the marketing problem.

(ii) It generates primary data for qualitative analysis.

Demerits of Pilot Studies

Some of the demerits of plot studies are as follows,

(i) The primary data comes from customers rather than knowledgeable persons.

(ii) Researchers cannot make decisions unless and until it is combined with secondary data.

II. Conclusive Research Design

It is a formal and a large research that involves quantitative analysis. The results of this study are used in decision-making and hence they are conclusive. It can be classified into descriptive research design and causative research design.

1. Descriptive Research Design

Descriptive studies are undertaken in many circumstances. When the researcher is interested in knowing the characteristics of certain groups such as age, sex, educational level, occupation or income, a descriptive study may be necessary.

Other cases, when a descriptive study could be taken up are when researcher is interested in knowing the proportion of people in a given population who have behaved in a particular manner, making projections of a certain thing or determining the relationship between two or more variables. The objective of such study is to answer the "who, what, when, where, and how" of the subject under investigation.

There is a general feeling that descriptive studies are factual and are very simple. But it may not necessarily be true. Descriptive studies can be complex, demanding a high degree of scientific skill on the part of the researcher.

Descriptive studies are well-structured. An exploratory study needs to be flexible in its approach, but a descriptive study, in contrast, tends to be rigid and its approach cannot be changed every now and then.

Types of Descriptive Research

Descriptive studies can be divided into two broad categories:

(i) Cross-Sectional and

(ii) Longitudinal.

Of the two, the former type of study is more frequently used.

(i) Cross-sectional Studies

A Cross-Sectional study is concerned with a sample of elements from a given population. Thus, it may deal with households, dealers, retail stores, or other entities. Data on a number of characteristics from the sample elements are collected and analyzed.

Cross-sectional studies are of two types: Field Studies and Surveys. Although the distinction between them is not clear, there are some practical differences which need different techniques and skills.

(a) Field Studies

Field studies are *ex-post-facto* scientific enquires that aim at finding the relations and interrelations among variables in a real setting. Such studies are done in life situations like communities, schools, factories, organizations, and institutions.

Field studies have their strengths and weaknesses. One major strength is that they are close to real life, and they cannot be criticized on the ground that they are remote from real settings or are artificial. Further, in real settings, variables exert their influence fully and, as such, the strength of variables is another advantage of field studies. Field studies are also strong in their heuristic quality.

Field studies are also subject to certain weaknesses. Such studies are scientifically inferior to laboratory and field experiments. One of their major weaknesses is their *ex-post- facto* character. As a result, interrelations among variables are weaker than they are in laboratory experiments. Another weakness is the lack of precision in the measurement of variables.

(b) Survey Research

Another type of cross-sectional study is Survey Research. A major strength of Survey Research is its wide scope. Detailed information can be obtained from a sample of a large population. Besides, it is economical as more information can be collected per unit of cost. Also, it is obvious that a sample survey needs less time than a census enquiry.

Despite these advantages of survey research, it is subject to certain limitations. Generally, survey research does not penetrate below the surface as more emphasis is given to the extent of information sought rather than to an in-depth analysis. Another disadvantage is that, Survey Research demands more time and more money, specially when it is conducted on a average scale. Another limitation of survey research is that the interview may makes the respondent alert and cautious and he may not answer the questions in a natural manner. Such answers will make the survey invalid.

(ii) Longitudinal Studies

Longitudinal studies are based on panel data and panel methods. A panel is a sample of respondents who are interviewed and then reinterviewed from time to time. Generally, panel data is related to the measurements of the same variables repeatedly.

Advantages of Longitudinal Studies

There are several advantages of using panel data,

1. Such data will enable the researcher to undertake detailed analysis.
2. Another advantage of the panel is that more comprehensive data could be obtained as individuals or families included in the panel are those who have accepted to provide data periodically. As panel members are willing persons, more data can be collected.
3. The other advantage is that panel data has been found to be more accurate than data collected through survey.
4. Costs of data collection through panels are generally lower than collected through personal interviews.

Limitations of Longitudinal Studies

There are certain limitations of panel data.

A major criticism of panels is that they may not be representative samples. This may distort the representative character of the original sample.

To minimize refusals of this type, many organizations pay some money to panel members. Usually, panel members are expected to act with a sense of responsibility and supply accurate information.

Moreover, after the initial attraction of membership of a panel has faded, members may loose interest in this task and may not fully cooperate with the research organization. This will affect the quality of information.

2. Causal Studies

Causal research investigates the cause and effect relationship between two or more variables. The design of causal research is based on reasoning along with tested lines. If it uses inductive logic for confirming or rejecting hypothesis with the help of further evidence. John Stuart Mill formulated a set of principles based on logic for causal research. The principles are: the Method of Agreement, the Method of Negative Agreement and the Method of Concomitant Variation.

The Method of Agreement states that, "When two or more cases of a given phenomenon have one and only one condition in common, then that condition may be regarded as the cause (or effect) of the phenomenon". Thus, if we find observation Z in every case where we find condition C we may conclude that C and Z are causally related.

This method helps in eliminating factors that are not relevant but also involve some limitations. One limitation is that, it assumes the variables that have not been considered are not relevant. One may not like to accept these assumptions as there are many variables considered but which may not influence Z, sometimes C operates only when certain conditions exist.

Finally, it may not be true that Z has occurred because of C. There may be a different cause for the occurrence of Z in each of the two situations. As against these limitations, it has some merits. First, it helps to rule out irrelevant factors, as a result of which the research problem becomes simpler. Second, it shows that certain factors do occur together. Third, it also shows that a particular factor, in this case C, always occurs before Z. Thus, it helps in eliminating extraneous factors, leaving factor C w'hich may be the cause.

Mill's second principle is the Negative Canon of agreement which states that when condition non-C is found to be associated with observation non-Z we may consider that a causal relationship exists between C and Z.

As the logical structures of both the methods are the same, the negative canon of agreement suffers from the same limitations as the first method. Here, also some other important factors might have been left out. It may be possible that non-C leads to non-Z only when some other factors are present.

Mill suggested another method, known as the Method of Concomitant Variation. This method holds that "If a change in the amount of one variable is accompanied by a comparable change in the amount of another variable, in two or more cases, and the latter change does not occur in the absence of the first change, then one change is the cause (or effect) of the other". This method indicates the process of correlation which may be either positive or negative. Again, each of these can be linear or non-linear.

Causal Interference Studies

Causal inference studies can be divided into two broad categories,

1. Natural experiments and
2. Controlled experiments.

1. Natural Experiments

Natural experiments are further divided into three. They are as follows,

(i) Time series and trend designs

(ii) Cross-sectional designs

(iii) A combination of the two.

2. Controlled Experiments

Controlled experiments are further divided into three. They are as follows,

(i) After only design

(ii) Before after with control group

(iii) Four group six study design etc.

2.4 Experimental Research Design

Q8. What is experimental design? Explain different types of experimental designs.

Ans : **(Nov.-20)**

Experimental design refers to the framework or structure of an experiment and as such there are several experimental designs. We can classify experimental designs into two broad categories, viz., informal experimental designs and formal experimental designs. Informal experimental designs are those designs that normally use a less sophisticated form of analysis based on differences in magnitudes, whereas formal experimental designs offer relatively more control and use precise statistical procedures for analysis. Important experiment designs are as follows;

1. **Before-and-after without Control Design:** In such a design a single test group or area is 'selected and the dependent variable is measured before the introduction of the treatment. The treatment is then introduced and the dependent variable is measured again after the treatment has been introduced. The effect of the treatment would be equal to the level of the phenomenon after the treatment minus the level of the phenomenon before the treatment. The design can be represented thus:

Test area:	Level of phenomenon before treatment (X)	Treatment introduced →	Level of phenomenon after treatment (Y)
	Treatment Effect = (Y) – (X)		

The main difficulty of such a design is that with the passage of time considerable extraneous variations may be there in its treatment effect.

2. **After-only with Control Design:** In this design two groups or areas (test area and control area) are selected and the treatment is introduced into the test area only. The dependent variable is then measured in both the areas at the same time. Treatment impact is assessed by subtracting the value of the dependent variable in the control area from its value in the test area. This can be exhibited in the following form:

Test area	Treatment introduced	Level of phenomenon after treatment (Y)
Control area:	⟶	Level of phenomenon without treatment (Z)
	Treatment Effect = (Y) – (Z)	

The basic assumption in such a design is that the two areas are identical with respect to their behaviour towards the phenomenon considered. If this assumption is not true, there is the possibility of extraneous variation entering into the treatment effect. However, data can be collected in such a design without the introduction of problems with the passage of time. In this respect the design is superior to before-and-after without control design.

3. **Before-and-after with control design:** In this design two areas are selected and the dependent variable is measured in both the areas for an identical time-period before the treatment. The treatment is then introduced into the test area only, and the dependent variable is measured in both for an identical time-period after the introduction of the treatment. The treatment effect is determined by subtracting the change in the dependent variable in the control area from the change in the dependent variable in test area. This design can be shown in this way:

	Time Period I		Time Period II
Test area:	Level of phenomenon before treatment (X)	Treatment introduced ⟶	Level of phenomenon after treatment (Y)
Control area:	Level of phenomenon without treatment (A)		Level of phenomenon without treatment (Z)
	Treatment Effect = (Y – X) – (Z – A)		

'This design is superior to the above two designs for the simple reason that it avoids extraneous variation resulting both from the passage of time and from non-comparability of the test and control areas. But at times, due to lack of historical data, time or a comparable control area, we should prefer to select one of the first two informal designs stated above.

3. **Factorial Designs:** Factorial designs are used in experiments where the effects of varying more than one factor are to be determined. They are specially important iri several economic and social phenomena where usually a large number of factors affect a particular problem. Factorial designs can be of two types:

   (i) Simple factorial designs and

   (ii) Complex factorial designs.

   (i) Simple factorial designs : In case of simple factorial designs, we consider the effects of varying two factors on the dependent variable, but when an experiment is done with more than two factors, we use complex factorial designs. Simple factorial design is also termed as a 'two-factor-factorial design', whereas complex factorial design is known as 'multi- factor-factorial design.'

2 × 2 Simple Factorial Design

Control variables	Experimental Variable: Treatment A	Experimental Variable: Treatment B
Level I	Cell 1	Cell 3
Level II	Cell 2	Cell 4

(ii) Complex factorial designs: Experiments with more than two factors at a time involve the use of complex factorial designs. A design which considers three or more independent variables simultaneously is called a complex factorial design. In case of three factors with one experimental variable having two treatments and two control variables, each one of which having two levels, the design used will be termed 2 × 2 × 2 complex factorial design which will contain a total of eight cells as shown below in Figure.

		Experimental Variable			
		Treatment A		Treatment B	
		Control Variable 2 Level I	Control Variable 2 Level II	Control Variable 2 Level I	Control Variable 2 Level II
Control Variable 1	Level I	Cell 1	Cell 3	Cell 5	Cell 7
	Level II	Cell 2	Cell 4	Cell 6	Cell 8

2.5 Research Design for Action Research

Q9. Discuss briefly action research and its process.

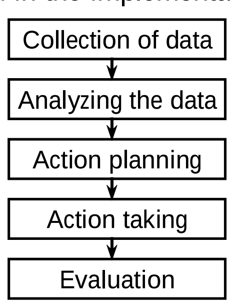
Ans :

Action Research

The concept of action research was developed by Kurt Lewin. It is defined as a "Data based, problem solving method which recreates the steps involved in the scientific inquiry method". Action research is mostly suited for planned changed programs. Action research integrates learning and doing. The dynamics of organizational change are learned and change efforts are implemented.

Action Research Process

The following are the steps involved in the implementation of an action research model,



Step 1: Collection of Data

After an action researcher identifies the problem the reasons, symptoms and results of the problems should be identified and accordingly the data should be collected for understanding the underlying structure of the problem. In this step, an action researcher defines the problem and information which need to be collected for solving ;the problem.

Step 2: Analyzing the Data

This is second step of action research process in which the data that is collected is analyzed for interpreting its meaning. An effective understanding and accurate interpretation of data helps in changing and improving the organization.

Step 3: Action Planning

This is one of the, most important step of action research process as it involves creating an action plans and assessing its potential impact. After deciding the course of action, the action steps should be outlined.

Step 4: Action Taking

In this step, an action is taken for solving the problem. This step is also called as intervention stage. Action researcher plays a very important role in this step by predicting unexpected results of the actions taken and planning for them.

Step 5: Evaluation

This is the final step of action research process in which the problems are appraised again. The current situation of the problem is analyzed and effect of action is evaluated. If necessary, corrections are also made in between the process.

Thus, action research process helps the individuals as well as organizations to learn, solve their problems and change their client systems.

2.6 Characteristics of the Good Research Design

Q10. What are the Characteristics of the Good Research Design?

Ans : **(Jan.-20, Imp.)**

Generally a good research design minimizes bias and maximizes the reliability of the data collected and analyzed. The design which gives the smallest experimental error is reported to be the best design in scientific investigation. Similarly, a design which yields maximum information and provides an opportunity for considering different aspects of a problem is considered to be the most appropriate and efficient design. A good research design possesses the following characteristics;

(i) Objectivity

It refers to the findings related to the method of data collection and scoring of the responses. The research design should permit the measuring instruments which are fairly objective in which every observer or judge scoring the performance must precisely give the same report. In other words, the objectivity of the procedure may be judged by the degree of agreement between the final scores assigned to different individuals by more than one independent observer. This ensures the objectivity of the collected data which shall be capable of analysis and interpretation.

(ii) Reliability

It refers to consistency throughout a series of measurements. For example, if a respondent gives out a response to a particular item, he is expected to give the same response to that item even if he is asked repeatedly. If he is changing his response to the same item, the consistency will be lost. So the researcher should frame the items in a questionnaire in such a way that it provides consistency or reliability.

(iii) Validity

Any measuring device or instrument is said to be valid when it measures what it is expected to measure. For example, an intelligence test conducted for measuring the IQ should measure only the intelligence and nothing else and the questionnaire shall be framed accordingly.

(iv) Generalizability

It means how best the data collected from the samples can be utilized for drawing certain generalizations applicable to a large group from which sample is drawn. Thus a research design helps an investigator to generalize his findings provided he has taken due care in defining the population, selecting the sample, deriving appropriate statistical analysis etc. while preparing the research design.

Thus a good research design is one which is methodologically prepared and should ensure that generalization is possible. For ensuring the generalization we should confirm that our research problem has the following characteristics;

(a) The problem is clearly formulated.

(b) The population is clearly defined.

(c) Most appropriate techniques of sample selection are used to form an appropriate sample.

(d) Appropriate statistical analysis has been carried out.

(e) The findings of the study are capable of generalizations.

(v) Adequate Information

The most important requirement of good research design is that it should provide adequate information so that the research problem can be analyzed on a wide perspective. An ideal design should take into account important factors like;

(i) Identifying the exact research problem to be studied

(ii) The objective of the research

(iii) The process of obtaining information

(iv) The availability of adequate and skilled manpower and

(v) The availability of adequate financial resources for carrying research.

(vi) Other Features

Some other important features of a good research design are flexibility, adaptability, efficiency, being economic and so on. A good research design should minimize bias and maximize reliability and generalization.

2.7 Data Collection Methods & Tools

Q11. Define data.

Ans :

Data refers to the information which is collected with the help of research for the purpose of making appropriate decisions. 'Morgenstem' quoted data in Jan. 1967 as,

"Data in itself is silent, it is the use to which it is put, in terms of inferring, interpreting, projecting, analysing, manipulating, computing and decision making that is important". Usually, only data is not useful. The collected data would be beneficial only if it provides good and reliable information. A researcher must use some strategy or method to provide a meaning to the collected data.

Q12. Explain briefly about collection of data?

Ans :

Data collection is the process to gather information about the relevant topic of research, which is being done by researcher. Data collection is a term used to describe a process of preparing and collecting data, for example, as part of a process improvement or similar project.

The purpose of data collection is to obtain information to keep on record, to make decisions about important issues, to pass information on to others. Primarily, data are collected to provide information regarding a specific topic.

Data collection usually takes place early in an improvement project, and is often formalized through data collection plan which often contains the following activity :

1. Pre collection activity on goals, target data, definitions and methods.
2. Collection of data
3. Presenting findings involving some form of sorting analysis and/or presentation.

While deciding about the method of data collection to be used for the study the researcher should keep in mind two types of data viz., primary data and secondary data. The primary data are those, which are collected afresh and for first time and thus happen to be original in character.

The secondary data are those which have been collected by someone else and which have already been passed through statistical process. The researcher would have to decide which sort of data he would be using for his study. The methods for collecting primary and secondary data differ since primary data are to be originally collected while in case of secondary data the nature of data collection work is merely that of compilation.

Database

Database systems have become an important tool in many applications over the past thirty years, ranging from traditional ones in business applications to more recent ones in science and technology. A database system is one of the critical components to build business applications. It provides an array of features which can be used to ensure optimal utilization of data for enhancing decision effectiveness in organizations.

The management of data in a database system is done by means of a general-purpose software package called a database management system. The database management system is the major software component of a database system. The DBMS helps to create an environment in which end user have better access to more and better managed data than they did before the DBMS become the data management standard.

A database is an organized list of facts and information. Database usually contain text and numbers, and frequently they hold still images, sounds and video or film clips. There is a difference between a simple list and a database? A database permits its user to extract a specific group of disparate facts from within a collection of facts. The paper filling system in an office is a kind of database. However, today people are more interested in databases constructed on computers where database management programs help people design and build collections of information.

Many databases offer citations and abstracts of journal articles and books. By searching for keywords that might appear in an article or book, a user can retrieve the citation, and often an abstract, of the journal article or book.

Uses of Database in Data Collection

Database can be used in following manner in data collection :

1. Citations and Abstracts

Many databases offer citations and abstracts of journal articles and books. By searching for keywords that might appear in an article or book, a user can retrieve the citation, and often an abstract, of the journal article or book.

2. Full text searching

A full-test database incorporates large files of text such as all of the paragraphs from a journal article or all of the chapters of a book. By searching for keywords that might appear in a text, a user can call forth the citation and abstract of a journal article or book. Today, many such databases do not offer the full text for free. Those with full text usually offer the search for citation and abstract for free, but charge for delivery of the full text of an article.

3. Updating

A typical database si designed around a central set of facts. All databases permit their operators to add new information or to update old facts whenever needed.

4. Sorting

In a database stored in a computer, the order of information can be rearranged or sorted quickly. To help people retrieve and print facts, databases sort lists into reports.

5. Electronic Storage

Most academic databases can be stored electronically, either on computers serving the Internet or on CD-ROM or DVD discs providing storage of research data.

6. Backup and Recovery

The system having this responsibility ensures that recovery is possible in the case of a system crash during execution of one or more transactions.

7. Redundancy can be Reduced

In a database system, the redundancy (duplicacy) of the data is controlled, and the system is aware of the redundancy and assumes the responsibility for propagating updates (automatic change at other places in case of duplicate data).

8. Inconsistency can be avoided

If the redundancy is controlled, the inconsistency is automatically controlled. Inconsistency means the mismatch in the two entries with same data. It must be avoided to get correct information.

2.7.1 Types of Data

Q13. What are the sources and methods of collecting data?

Ans :

To understand the multitude of choices available to a researcher for collecting the project/ study-specific information, one needs to be fully cognizant of the resources available for the study and the level of accuracy required. To appreciate the truth of this statement, one needs to examine the gamut of methods available to the researcher. The data sources could be either contextual and primary or historical and secondary in nature (Figure).

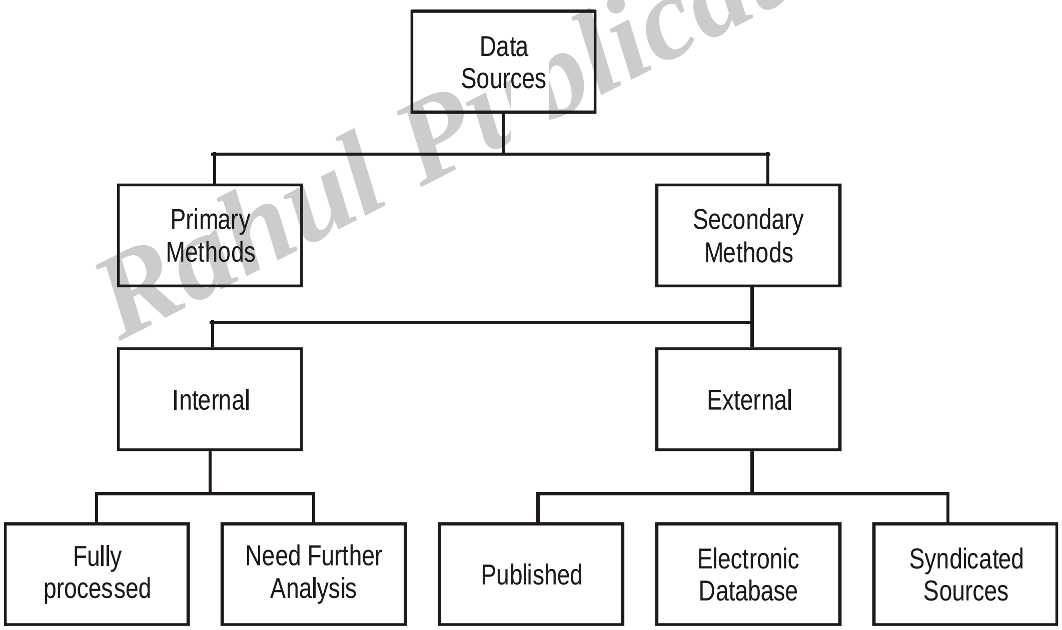


Fig.: Sources of research information

1. **Primary data** as the name suggests is original, problem- or project-specific and collected for the specific objectives and needs spelt out by the researcher. The authenticity and relevance is reasonably high. The monetary and resource implications of this are quite high and sometimes a researcher might not have the resources or the time or both to go ahead with this method. In this case, the researcher can look at alternative sources of data which are economical and authentic enough to take the study forward. These include the second category of data sources—namely the secondary data.

2. **Secondary data** as the name implies is that information which is not topical or research-specific and has been collected and compiled by some other researcher or investigative body. The said information is recorded and published in a structured format, and thus, is quicker to access and manage. Secondly, in most instances, unless it is a data product, it is not too expensive to collect. As suggested in the opening vignette, the data to track consumer preferences is readily available and the information required is readily available as a data product or as the audit information which the researcher or the organization can procure and use it for arriving at quick decisions.

   In comparison to the original research-centric data, secondary data can be economically and quickly collected by the decision maker in a short span of time. Also the information collected is contextual; what is primary and original for one researcher would essentially become secondary and historical for someone else.

2.7.2 Sources and Instruments for data

Q14. Explain the various methods of collecting primary data.

Ans : **(Jan.-20, Imp.)**

Primary data is one, which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by surveys conducted by individuals or research institutions. Primary data collection is necessary when a researcher cannot find the data needed in secondary sources.

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

Conducting primary research is a useful skill to acquire as it can greatly supplement research in secondary sources, such as journals, magazines, or books. Primary research is an excellent skill to learn as it can be useful in a variety of settings including business, personal, and academic.

Methods of Collecting Primary Data

The various methods of collecting primary data are :

(a) Interviews
   - Personal Interview
   - Telephone Interview

(b) Surveys

(c) Observations

(d) Analysis

(e) Questionnaires

(f) Schedules

(a) Interviews

Interviews are non-on-one or small group question and answer sessions. Interviews will provide a lot of information from a small number of people and are useful when you want to get an expert or knowledgeable opinion on a subject. The interviews can be :

- **Personal Interview :** The investigator follows a rigid procedure and seeks answers to a set of pre-conceived questions through personal interviews. This method of collecting data is usually carried out in a structured way where output depends upon the ability of the interviewer to a large extent.

- **Telephone Interviews :** This method of collecting information involves contracting the respondents on telephone itself. This is not a very widely used method but it plays an important role in industrial surveys in developed regions, particularly, when the survey has to be accomplished in a very limited time.

(b) Surveys

Surveys are a form of questioning that is more rigid than interviews and that involve larger groups of people. Surveys will provide a limited amount of information from a large group of people and are useful when you want to learn what a larger population thinks.

(c) Observations

Observations involve taking organized notes about occurrences in the world. Observations provide you insight about specific people, events, or locales and are useful when you want to learn more about an event without the biased viewpoint of an interview. The information obtained relates to what is currently happening and is not complicated by either the past behaviour or future intentions or attitudes or respondents.

This method is no doubt an expensive method and the information provided by this method is also very limited. As such this method is not suitable in inquiries where large samples are concerned.

(d) Analysis

Analysis involves collecting data and organizing it in some fashion based on criteria you develop. They are useful when you want to find some trend or pattern. A type of analysis would be to record commercials on three major television networks and analyze gender roles.

(e) Questionnaire

The questionnaire is an important tool for gathering primary data. Questionnaire are mailed to the respondents with a required to return after completing the same. It is most extensively used method in various economic and business surveys. Questionnaire to be used must be prepared very carefully so that it may prove to be effective in collecting the relevant information. Poorly constructed questions can result in large errors and invalidate the research data, so significant effort should be put into the questionnaire.

(f) Schedules

Under this method the enumerators are appointed and given training. They are provided with schedules containing relevant questions. These enumerators go to respondents with these schedules. Data are collected by filling up the schedules by enumerators on the basis of replies given by respondents. Much depends upon the capability of enumerators so far as this method is concerned. Some occasional field checks on the work of the enumerators may ensure sincere work.

Q15. What are the advantages and disadvantages of primary data.

Ans :

Advantages of Primary Data

1. Reliability

The information collected for primary data is more reliable than those collected from the secondary data because this information is collected directly from the respondents.

2. Availability of a Wide Range of Techniques

There are lot of techniques that can be employed, which means that all information necessary can be obtained by using the appropriate techniques, enabling all areas of the research topic to be answered and investigating thoroughly and effectively.

3. Addresses Specific Research Issues

The organisation asking for the research has the complete control on the process and the research is streamlines as far as its objectives and scope is concerned. Researching company can be asked to concentrate their efforts to find data regarding specific market rather than concentration on mass market. Primary research is designed to collect the information the marketer wants to know, and report it in ways that benefit the marketer.

4. Greater Control

Not only does primary research enable the marketer to focus on specific issues, it also enables the marketer to have a higher level of control over how the information is collected. In this way the marketer can decide on such issues as size of project (e.g., how many responses), location of research (e.g., geographic area) and time frame for completing the project.

5. Efficient Spending for Information

Unlike secondary research, where the marketer may spend for information that is not needed, primary data collection focuses on issues specific to the researcher and improves the chances that research funds will be spent efficiently.

6. **Proprietary Information**

Information collected by the marketer using primary research is their own and is generally not shared with others. Thus, information can be kept hidden from competitors, and potentially offer an "information advantage" to the company that undertook the primary research.

Disadvantages of Primary Data

1. **Cost**

Compared to secondary research, primary data may be very expensive since there is a great deal of marketer involvement and the expense in preparing and carrying-out research can be high. Skilful persons are required, which is also not an economic process, it is costly.

2. **Time Consuming**

To be done correctly primary data collection requires the development and execution of a research plan. Going from the start-point of deciding to undertake a research project to the end-point of having results is often much longer than the time it takes to acquire secondary data.

3. **Not Always Feasible**

Some research projects, while potentially offering information that could prove quite valuable, are not within the reach of a marketer. Many are just too large to be carried-out by all but the largest companies, and some are not feasible at all.

4. **Large Volume of Data**

Since the data collected by primary methods remains in a very large amount so it becomes very complicated to handle and maintain all the data. Large volumes of data also create difficulty in the data processing.

5. **Reluctancy of Respondents**

In many cases, the respondents remain reluctant to give the answers of the researchers' questions. Sometimes they give such answers which create business in the research.

Q16. Explain the various ways editing of primary data.

Ans :

1. **Editing for Completeness**

The editor should see that each schedule and questionnaire is complete in all respects, *i.e.,* answer to each and every question has been furnished. If some questions have not been answered and those questions are of vital importance the informants should be contacted again either personally or through correspondence. It may happen that in spite of best efforts a few questions remain unanswered. In such questions, the editor should mark "No answer" in the space provided for answers and if the questions are of vital importance then the schedule or questionnaire should be dropped.

2. **Editing for Consistency**

While editing the data for consistency, the editor should see that the answers to questions are not contradictory in nature. If there are mutually contradictory answers, he should try to obtain the correct answers either by referring back the questionnaire or by contacting, wherever possible, the informant in person. For example, if amongst others, two questions in questionnaire are: (a) Are you married? (b) State the number of children you have, and the reply to the former question is 'no' and to the latter 'three', then there is contradiction and it should be clarified.

3. **Editing for Accuracy**

The reliability of conclusions depends basically on the correctness of information. If the information supplied is wrong, conclusions can never be valid. It is, therefore, necessary for the editor to see that the information is accurate in all respects. However, this is one of the most difficult tasks of the editor. If the inaccuracy is due to arithmetical errors, it can be easily detected and corrected. But if the cause of inaccuracy is faulty information supplied, it may be difficult to verify it, *e.g.,* information relating to income, age, etc.

4. Editing for homogeneity

By homogeneity we mean the condition in which all the questions have been understood in the same sense. The editor must check all the questions for uniform interpretation. For example, as to the question of income, if some informants have given monthly income, others annual income and still others weekly income or even daily income, no comparison can be made. Similarly, if some persons have given the basic income whereas others the total income, no comparison is possible. The editor should check up that the information supplied by the various people is homogeneous and uniform.

Q17. Explain the various methods of collecting secondary data.

Ans : **(Nov.-21)**

Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency.

Secondary data is data that is neither collected directly by the user nor specifically for the user, often under conditions not known to the user like Government reports. Secondary information has already been collected for some other purposes. It may be available from internal sources, or may have been collected and published by another organisation. Secondary data is cheaper and more quickly available than primary data, but needs processing before it is useful. Since the data has been collected for another purpose by somebody else, it may not be fully useful, the context could have changed or data could have been doctored.

According to W.A. Neiswanger, "A secondary source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible".

The purpose of collecting secondary data is to make some changes, or to review the needs, to make solutions based on normal database management. To get some new ideas or it was also for the purpose of time saving. If researcher has short time and researcher has to complete an objective, secondary data in this regard is the best way to save time and complete task.

Tools and Techniques of Collecting Secondary Data

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

1. Internal Secondary Data

Data that originate within the firm for which the research is being conducted are internal data. They may be adapted for the marketing research purposes. They may be formal data and informal data. Formal data are available on a regularly scheduled basis, such as monthly, quarterly or annually in a form that allows comparisons through time. Informal data reports basic marketing knowledge and are available on a non-recurring basis.

Sales analysis and invoicing are considered important sources of internal secondary data.

i) Sales Analysis

Sales analysis is an important tool of marketing research. It is the first step in the marketing research program and acts as a basis for the development of further marketing research. It reveals the current operating problems in the marketing area where the scope for marketing research can be adequately explored in smaller organisations, sales analysis is an important source of marketing information. It provides a major share of the factors for marketing research.

ii) Invoice Analysis

Company invoices have been a very useful source of information. A copy of an invoice is preserved and information from it may be punched, tabulated, processed and summarized to provide suitable information to the researcher. The invoice data may be classified according to customer, nature of product, region and area. The invoice record may be of immense use provided it has been used with precaution and scientifically.

iii) Accounting Records

The basis for accounting records concerned with sales is the sales invoice. The usual sales invoice has a sizable amount of information on it, which generally includes name of customer, location of customer, items ordered, quantities ordered, quantities shipped, dollar extensions, back orders, discounts allowed and date. In addition, the invoice often contains information on sales territory, sales representative and warehouse of shipment. This information, when supplemented by data on costs and industry and product classification, as well as from sales calls, provides the basis for a comprehensive analysis of sales by product, customer, industry, geographic area, sales territory, and sales representative, as well as the profitability of each sales category. Unfortunately, most firms' accounting systems are designed primarily for tax reasons rather than for decision support.

2. External Secondary Data

The second form of secondary data is external sources which are generally published and are available in different forms and from different sources. Although external secondary data may be obtained from different sources, some of the sources are given here.

i) Libraries

Researchers first attend libraries to find out relevant data pertaining to research. They provide many sources where suitable data may be obtained. Public libraries, colleges and University libraries contain a large amount of business information, which provides sources of other data. Management books, theses, management journals and other publications can be consulted in these libraries. Management institutes, research institutes, banks, insurance companies, public utility companies and manufacturing units have maintained adequate libraries.

ii) Literature

A great amount of secondary data is available from literature, particularly on marketing subjects. With the development of marketing researches in different countries, new and interesting facts are coming into the picture, which are available in various publications. Consultations of this literature may provide proper guidance pertaining to publication, which can be used from time to time.

iii) Periodicals

Business periodicals published fortnightly, monthly, quarterly, semi annually and annually are often consulted by the marketing executives and researchers to plan and design their marketing research. Periodicals on economics, finance, trade, transport, industry, labour and management are being prepared by the Government as well as by the non-government agencies. Journals of the marketing association, management association, research agencies, advertising agencies and other related periodicals are becoming very common in India and abroad.

iv) Census and Registration Data

Census and registration data have become very comprehensive sources of marketing research. Previously, these were concentrated only to one population census, but it now extends too many areas. Census and registration data includes Census of Population, Census of Agriculture, Census of Cattle, Census of Trade, Census of Transport, Census of Industry, Census of Banking and Finance etc.

v) Trade Associations

Trade associations may be an excellent source of data pertaining to an industry. The trade association of one industry may exchange data with the trade association of another industry, and within one industry a firm may exchange data with another firm with the help of trade association of the industry.

vi) Government Departments

Different government departments have different data, which are not available in libraries. But these are very useful for understanding various aspects of the economy. The researchers can utilize them for the purposes of their research. Information and data pertaining to agriculture, industry, trade, transport, banking and finance can be obtained from the respective ministries of the Government of India.

vii) Private Sources

Private sources include varied sources available in the form of books, monographs, bulletins, journals, commercial reports and so on. They are priced and publicly circulated. Some of the sources include extensive original research and some summarise the research findings of other person. Many of them are statements of facts and opinions. The All India Management Association, the Indian Marketing Association, Commerce Pvt. Ltd., Capital, the Economic Times and Financial Express etc. are important private institutions which supply suitable information and data to the public in the form of journals, books and newspapers.

viii) Commercial Data

There are several institutions and companies, which purchase and sell marketing information and data. Some of these companies are solely engaged in marketing research. They collect information and data directly from field surveys. Some such companies collect and process the secondary data and supply them to their subscribers.

ix) Financial Data

The financial data of reputed concerns are available in several magazines, newspapers, journals and in summary of statistics. The Directorate of Income Tax publishes information pertaining to taxes and income ranges. Such information and data are useful to forecast the market potential of a particular product. Private institutions such as the Economic Times of India, Commerce Private Ltd. etc. are publishing assets and investment-wise data of several large companies. The market researchers are indirectly benefited by such data and information.

x) International Organisations

International organisations such as the International Monetary Fund, the World Bank, the United Nations Organisation, the Asian Bank, the African Bank, Foreign embassies etc. publish several useful statistics, which can be used by researchers. The statistics may relate to the population problem, trade, institutions, culture, agriculture, regional festivals, superstitions, education, consumption, transportation, forestry, manufacturing and so on. There are several other publications such as the World Almanac, Thomas Register etc., which publish much useful information for marketing researchers.

xi) References and Bibliography

In every publication, the researcher can find references and a bibliography which can be very good sources of information of marketing research. The researcher can consult them for further information and data.

xii) Volumes of Statistics

There are several private and public organisations, which prepare a summary of statistics. In India, the Indian Statistical Institute publishes the Statistical Abstract. Commerce Pvt. Ltd., the Times of India Ltd. and the Financial Express compile directories of different subjects. The Government of India publishes the economic survey of India wherein statistics relating to every field of economic activities are compiled in a suitable form.

xiii) Advertising Agencies

Advertising agencies have proved to be very useful sources of marketing research. Recently, a large number of agencies have come into the findings of the advertising researches for their clients. Advertising agencies sometimes, publish reviews, resumes and tests of marketing researches. The consumers' behaviour, consumption pattern and demographic features are generally revealed by these agencies.

xiv) Other Sources

There are several other sources of marketing researches. Individuals conduct their own researches, which may be purchased by other institutions. Marketing associations, management associations and individual business houses have been conducting marketing researches for other researches. There are a large number of researches organisations in foreign countries, which are selling their research findings to organisations requiring knowing the outcome of their researches.

Q18. What are the advantages and disadvantages of Secondary Data?

Ans :

Advantages of Secondary Data

1. Economy

Such data is cheaper. The amount of money spent in acquiring secondary data is generally less than that needed to obtain primary data. The various secondary data from libraries can be obtained at no cost.

2. Quickness

Most of these data are 'instant' since they already exist and merely need to be discovered. Thus, the time in collecting secondary data is largely search time and usually requires few hours or days.

3. Quality

An individual investigator cannot match the quality or size of the firms that obtain much of the existing secondary data. The information is gathered by trained personnel specialised in data collection. Also many organisations may not release their data to individual researcher but may give it to firms.

4. No Need of Measuring Instruments

When information is gathered from secondary sources, there is a problem in designing information gathering instruments as information is already collected by someone else. The only problem is to locate the appropriate source and method of recording desired information.

5. Availability

Secondary data is sometimes available even in those cases where primary investigations are not possible.

6. Bases for Comparison

Secondary data is useful in the case of exploratory researches as they provide increased understanding of the problem.

7. Useful in Exploratory Research

Secondary data act as a basis for comparison after primary data is collected.

8. Generates Feasible Alternatives

Secondary data are very useful in generating viable alternatives to solve problems. The multiplicity of data sources, research approaches and managerial styles usually lead to a number of possibilities which should be examined by the researcher.

Disadvantages of Secondary Data

It is however difficult to find secondary data that exactly fits into the needs of some specific research investigation. The problems experienced in respect of secondary sources can be in terms of :

1. Relevance

The data may not fit into the needs of investigation. There may be difference in the units of measurement; there may be surrogated data; discrepancy of classes and data may pertain to some other period of time.

2. Accuracy

It is observed that it is rather difficult to measure the degree of approximations used in the collection of information as well as the competence of the investigator in motivating the persons to supply the desired information.

3. Existence of Obsolete Information

Information may be outdated or obsolete.

4. Non-Disclosure of Research Findings

All the findings of a research study may not be made public.

5. Seldom Catering to the Need

The available data may not suit the current purpose of research, due to incompleteness, generalities and so on.

6. Other Limitations

i) There may be difficulties in the identification of the source.

ii) Errors may be there in recording or transferring information from secondary sources.

iii) The facilities or capabilities of the agency that originally collect the data might be questionable.

Q19. Explain the various ways editing of secondary data.

Ans :

Since secondary data have already been obtained it is highly desirable that a proper scrutiny of such data is made before they are used by the investigator. In fact, the user has to be extra-cautious while using secondary data. In this context, Prof. Bowley rightly points out that secondary data should not be accepted at their face value." The reason is that such data may be erroneous in many respects due to bias, inadequate size of the sample, substitution, errors of definition, arithmetical errors, etc. Even if there is no error such data may not be suitable and adequate for the purpose of the enquiry. Hence Before using such data, the investigator should consider the following aspects:

Whether the data are suitable for the Purpose of Investigation in View

Before using secondary data the investigator must ensure that the data are suitable for the purpose of the equity. The suitability of data can be judged in the light of the nature and scope of investigation. For example, if the object of enquiry is to study the wage levels including allowances of workers and the data relate to basic wages alone, such data would not be suitable for the immediate purpose. It may be difficult to find data which exactly fit the needs of the present project.

Quite often secondary data do not satisfy immediate needs because they have been compiled for other purpose. Even when directly pertinent to the subject under study, secondary data may be just enough off the point to make them of little or no use. The value of secondary data is frequently impaired by:

(i) Variation in the units of measurement: consumer income, for example, may be measured by individual, family household, spending units or tax return.

(ii) Definition of classes may be different: for example, definition of literate, educated, poor may vary from researcher to researcher.

(iii) Variation in the date/period to which the data related: the data available may relate to a different time period and not serve the purpose of researcher. Data published to promote the interests of a particular group whether it is political, social or commercial are suspect.

Whether the Data are Adequate for the Investigation If it is found that the data are suitable for the purpose of investigation, they should be tested for adequacy. Adequacy of the data is to be judged in the light of the requirements of the survey and the geographical area covered by the available data. For example, in the illustration given above, if our object is to study the wage rates of the workers in sugar industry in India and if the available data cover only the State of U.P., it would not serve the purpose. The question of adequacy may also be considered in the light of the time period for which the data are available. For example, for studying trend of prices we may use data for the last 8-10 years but from the source known to us data may be available for the last 2-3 years only which would not serve the purpose.

Whether the Data are Reliable It is very difficult to find out whether the secondary data are reliable or not. The following tests, if applied, may be helpful to determine how far the given data are reliable:

(i) Which specific method of data collection was used? If a source fails to give a detailed description of its method of data collection, researchers should be hesitant about using the information provided. When the methodology is described, researchers should subject it to a painstaking examination.

Data published to promote the interests of a particular group whether it is political, commercial or social are suspect.

(ii) Was the collecting agency unbiased or did it "have an axe to grind"?

(iii) If the enumeration was based on a sample, was the sample representative?

(vi) Were the enumerators capable and properly trained? Incompetent or poorly trained enumerators cannot be depended upon to produce useful result.

(v) Was there a proper check on the accuracy of field work?

(vi) Was the editing, tabulating and analysis carefully and conscientiously done? Carelessness in either one or more of these functions can render of little value the findings of an otherwise valuable study.

(vii) What degrees of accuracy was desired by the compiler? How far was it achieved?

Q20. What are the difference between Primary and Secondary Data ?

Ans : **(Imp.)**

Basis of Difference	Primary Data	Secondary Data
1) **Meaning**	Primary data is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study.	Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency.
2) **Cost**	The cost of obtaining primary data is typically more.	Secondary data is typically available for free or for the subscription fee to the database magazine or journal.
3) **Sources**	A primary data source is a publication in which the data are published by the same authority which gathered and analysed them.	A secondary data source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible.
4) **Methods**	It is normally collected through experiments, surveys, questionnaire focus groups, interviews.	The method or way of collecting secondary data includes books, journals, census data biographies, articles and databases.
5) **Reliability**	The information collected for primary data is more reliable than those collected from the secondary data.	Whereas secondary data are less reliable because these information are not collected for that particular purpose.
6) **Scientific Method**	Primary data follows the scientific method. A hypothesis is formed, data is collected from an experiment based on the hypothesis and the hypothesis is proven correct or not.	Secondary data does not start with a hypothesis as the data is already collected. Patterns and insights are found within the secondary data and then the observation on that data is made.
7) **Precaution**	No extra precautions are required in primary data.	Secondary data need more care and attention.
8) **Form of Data**	Primary data are in the shape of raw material.	Secondary data are usually in the shape of readymade products.
9) **Accuracy**	Primary research is tailored specifically for the project and tends to be more accurate. Primary data is customised.	While secondary data can provide plenty of information, it is less accurate because the data collected was not collected specifically for the questions. Secondary data is not customised.
10) **Example**	One's own questionnaire.	Data from a magazine, Journal, etc.

2.8 Questionnaire

Q21. Define Questionnaire. Explain different types of Questionnaire.

Ans : **(Imp.)**

Questionnaire is a data collection instrument. A questionnaire is a prepared set of questions (or measures) used by respondents or interviewers to record answers (data). The term questionnaire usually refers to a self- administered process where by the respondent himself reads the question and records his answers without the assistance of an interviewer. This is a narrow definition of a questionnaire.

A questionnaire is a method of obtaining specific information about a defined problem so that the data, after analysis and interpretation, results in a better appreciation of the problem. A questionnaire form, which has to be completed by an interviewer, is often referred as schedule.

The success of collecting data either through the questionnaire method or through the schedule method depends largely on the proper design of the questionnaire. This is a specialised job and requires high degree of skill, experience, through knowledge of the research topic, ability to frame questions and a great deal of patience. There are no hard and fast rules in designing the questionnaire.

Questionnaire design is only one phase of several interrelated business research steps. But it is a very important phase because data collected with questionnaires is used to improve decision-making. Moreover, it is a structured framework consisting of a set of questions and scales designed to generate primary data. When designing a questionnaire, researchers must realise that there will be only one opportunity to interact with respondents, since a reasonable interval of time is necessary before the same respondent can be contacted again, and then it should generally involve either another topic or a different approach to the same topic. A questionnaire is quantitative research, in that it can provide lots of responses relatively easily. It is a good way of asking a large number of people straightforward questions.

Types of Questionnaires

Questionnaires can be constructed so that the objective is clear to the respondent (none disguised); or they can be constructed so as to disguise the objective. Using these two bases of classifications, four types of studies can be distinguished as below:

1. **Structured, Non-Disguised Questionnaire**

   Most questionnaire studies made in marketing research are of the first type- they are structured and are not disguised. If the sales manager for a musical instrument company wants to find out how many and what type of people play various types of instruments, a formal list of questions may be set up that asks directly about the ownership and playing of various instruments. Each of a selected group of persons is then asked this set of questions in the given sequence. Answers are frequently limited to a list of alternatives, which is stated or implied. Several questions taken from an actual survey of this type are given below.

   Structured, non-disguised studies can be handled by telephone, mail, or personal interview. They are subject to the three limitations of the questionnaire -method- respondents may be unable to furnish the information desired, they may be unwilling to furnish it, or the questioning process may tend to stimulate incorrect or misleading answers.

2. **Non-Structured, Non-Disguised Questionnaire**

   The purpose of the study is clear, but the responses to the question are open-ended. For 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 very 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 respondent fairly and accurately. The main disadvantage of this method of interviewing is that it takes time, and the respondents may not cooperate. Another disadvantage is that coding of open-ended questions may pose a challenge.

To overcome these difficulties, researchers have developed depth interviews and focus-group interviews. Instead of approaching respondents with a fixed list of questions, the interviewer attempts to get respondents to talk freely about the subject of interest.

3. Non-Structured, 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 :

i) Word association,

ii) Sentence completion, and

iii) Storytelling.

4. Structured, Disguised Questionnaire

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 they feels. Therefore, the endeavour in this method is to know the respondent's attitude.

Some structured, disguised tests- of attitudes are based on the theory that individuals' knowledge, perception, and memory are conditioned by their attitudes. For example, "What do you think about the Babri Masjid demolition?"

Q22. What are the Components of Questionnaire.

Ans :

1. Words

The most obvious component is words. Researchers must carefully consider which words to use in creating the questions and scales for collecting raw data from respondents. The words selected by the researcher can influence respondent's answers to a given question. A few example of wording problem include ambiguity, abstraction and connotation.

2. Questions/Setups

The next component is question setup used in particular scale to collect raw data from the respondents. Two important issue relating to question phrasing that have a direct impact on survey design are the quality of question and the type of question format :

i) Simple Alternative Questions: Such questions can be answered in 'yes' or 'no' or 'right' or 'wrong'.

ii) Multiple Choice Questions: Such questions may be answered in a number of ways. The answers should be printed in the questionnaire itself, and the informant should be requested to mark against any one of them.

iii) Specific Information Questions: Such questions solicit specific information like, - What is your age?

iv) Open Questions: Such questions are to be answered by the informants in their own words.

3. Questionnaire Format

This component does not directly relate to the process of developing the individual questions but rather the layout of sets of questions or scale measurements into a systematic instrument. The questionnaire's format should allow for clear communication.

4. Hypothesis Development

The final component focuses on the notion that questionnaires are designed for collecting meaningful data to test a hypothesis rather than merely to gather facts. Theoretically, each of the components should either directly or indirectly relate to a research hypothesis that is relevant to the research objectives. Hypothesis can relate to following things :

i) The nature of the respondent.

ii) The relationship between expressed attitude and behaviour of the respondents (e.g. motivation)

iii) The sociological structures and their influence on the respondent.

iv) The meaning of words and the respondent's grasp of language and/or concepts.

v) The relationships among a respondent's knowledge, attitude and marketplace behaviours.

vi) The descriptive and productive capabilities of attributes of the constructs.

Q23. How to Organizing / Designing of Questionnaire?

Ans :

Adequate preparation of questionnaire or questionnaire designing is critical to the success of a survey.

Questionnaire designing is discussed in following steps. These steps may vary in importance in individual projects, but each step must receive attention in each case. The steps of constructing questionnaire are :

1. Determine What Information is Wanted

Questionnaires are prepared to meet research objectives and to motivate the respondents to cooperate with the survey. Therefore a specific statement of the information required for research purposes is prepared and put in operation to motivate the respondents. The specific characteristics of the information are decided upon for the proposed analysis and objectives.

2. Determine the Type of Questionnaire to Use

After deciding the information required for the research, the next step is to decide the method of using the questionnaire or administering the questionnaire. The questionnaire can be used by personal interview, mail, telephone or all of them.

The choice among these alternatives is largely determined by the type of information to be obtained and by the type of respondents from whom it is to be obtained. It is necessary to decide on the type of questionnaire at this point since the questions asked, the way in which they are asked, and the sequence in which they are asked will all be influenced by this decision. The influence of the type of questionnaire on these factors will be brought out in the discussion.

3. Determine the Content of Individual Questions

Once the needed information is specified, the method of communication is decided, researchers are ready to begin formulating the questionnaire. One problem is to decide what to include in individual questions.

4. Determine the Type of Question to Use

Once the content of individual questions is decided, researchers are ready to begin forming the actual questions. Before they can work on the wording of each question, they must decide on the type of question to use. Part of this decision is whether to use disguised or non disguised, structured or unstructured questioning.

5. Deciding on Wording of Questions

In the preceding discussion of question content and types of questions, much has been said on question wording. A number of other important ideas however should be considered. Unfortunately, these ideas are more rules of thumb that have been developed form experience than they are underlying concept.

i) Define the issue.

ii) Should question be subjective or objective?

iii) Positive or negative statement.

iv) Use simple words.

v) Avoid ambiguous questions.

vi) Avoid leading questions.

6. Decide on Question Sequence

Once the wording of the individual questions has been determined, it is necessary to set them up in some order. The sequence can influence the results obtained. A questionnaire has three major sections:

i) Basic information

ii) Classification information

iii) Identification information

7. Decide on Length of Questionnaire

How long the questionnaire/ schedule would be defends upon:

i) What the researcher wants to know and how many items are necessary so that the data will be credible;

ii) On the type of study (since self-administered questionnaires may be shorter than face-to-face interviews;

iii) On the time which the researcher has available for the study;

iv) On the time the respondents can and will take; and

v) On researcher's resources.

For obtaining necessary and adequate data and credible answers, it is necessary that the length of the questionnaire should be given importance, i.e., it should be reasonably long. It is equally important that time for filling up questionnaires or responding to interview schedule is generally limited to 30-40 minutes in comparison to face-to-face interview which can continue for 45-60 minutes. Another consideration is the respondents. How long can they be available? Will they take interest in answering questions seriously? Young people may be available for less time than the middle-aged and the old people.

8. Decide on Layout and Reproduction

The physical layout and reproduction of the questionnaire influence the success of the interview. While planning the layout and reproduction, three important points are considered. They are :

i) **Acceptance of the Question- naire:** The physical appearance of the questionnaire influences the interests and attitude of the respondents. If the questionnaire is prepared on rough paper, typed unimpressively and designed poorly it may not attract the respondents to read and answer the questions. On the other hand, a questionnaire typed or printed on good quality paper may attract the respondents to read it. The respondents may be requested not to disclose their identities. The researchers, sometimes, to avoid any bias, may avoid the use of their company's name.

ii) **Ease of Control:** The questionnaire should be numbered serially to make it possible to control the questionnaire in the operation. It will make it easy to edit and tabulate the answers. All questions must be accounted and evaluated properly. Numbered questions are easy to follow and simple to operate throughout the survey time, and analyse thereafter.

iii) **Ease of Handling:** The reproduction of a questionnaire may influence the fieldwork and analysis. It is essential that a large number of questions must not be put in a short space. If the questionnaire is crowded, it makes a bad appearance. This may cause errors in collection of data and tabulation as it is hard to read the answers. Too large a questionnaire cannot be handled properly. Questions should be laid out and reproduced in an easy way for the field worker to follow the sequence.

9. Check Questions

Once the first draft of the questionnaire has been completed, and before it is actually pre-tested, it is a good idea to get one (or more) expert's opinion of the questionnaire. A person who is expert in research methodology can help to catch methodological weaknesses in the instrument, such as faulty scales, inadequate instructions, etc. A person who is familiar with the topic of the questionnaire can help in assessing the face validity of the questions. Do they make sense; are they easy to understand, do they ask what they are supposed to be asking?

10. Pilot-Testing or Pre-Testing

Before the questionnaire is ready for the field, it needs to be pre-tested under field conditions. No researcher can prepare a questionnaire so good that improvements cannot be discovered in field test. Researchers have reported pre-testing, changing, and pre-testing again for as many as 25 times before they were satisfied with some questionnaires. One pre-test is as much, however, as most questionnaires get.

11. Revision and Final Draft

After each significant revision of the questionnaire, another pre-test should be run. When the last pre-test suggests no new revisions, the researcher is ready to print the actual questionnaires to be used in the survey.

2.8.1 Guidelines for Questionnaire

Q24. What are the Guidelines / Precautions for Preparation of Questionnaire ?

Ans : **(Imp.)**

The guidelines for preparing questionnaire are as follows :

1. Arrange Questions in a Logical Order

Arrange questions carefully so that the respondents will be able to make their replies easily and without confusion. Ask an easy-to-answer question in the beginning. Also, group the sequence of items in a logical and coherent order. If possible, group together all items about a particular topic or subject. This grouping will help the respondent think more logically about the issues involved. It will show the thoughtful plan in designing the questionnaire.

2. Design Items that Require Current and Easily Remembered Data

When respondents must rely too much on their memories they may either guess or not respond at all. Either way, their answers would be invalid and unreliable.

3. Questions should not be Ambiguous

The structure of the sentence as well as the word choice and order should not provide any room for misinterpretation on the part of the respondent. A question like 'What type of T.V. do you prefer?' is ambiguous. The respondent may get a doubt - Do they mean coloured or black and white? Or portable or non-portable? Or perhaps it is the brand they are after. The question could be much more accurately phrased in such a manner as this; 'Do you prefer portable or non-portable T.V.?'

4. Leading Questions should not be Asked

These are questions that suggest the desired answer or anticipate answers. These questions condition the respondent's mind. So the respondent cannot give the truthful answer. For example, 'Do you read The Economic Times?' is a leading question. A better question to ask would be - 'Which newspaper do you read?'

5. Personal Questions should be Avoided

Theses include question about politics, religion, age and income, etc. Sometimes one needs the information generally. For example, you will not need the exact age or exact amount of income. However, the information may be necessary in order to accomplish the particular purpose of the report. To encourage response as well as facilitate evaluation of the answers, provide ranges from which the respondent may show his age and income range. For example, the researcher can show age brackets below 21, '20 to 25' and set-up income brackets in the same way.

6. Good Transition between Questions

Provide good transition between questions and if possible use parallel wording. Both of these factors will aid the respondent in moving from one question to another. With good transition he easily sees the connection between questions. Parallel wording actually makes it easier for him to understand questions and thus answer them.

7. Avoid Skip-and-Jump or Involved Rating Questions

If possible, 'skip-and-jump' or 'involved rating' questions should not be asked. For the average person the 'skip-and-jump' type of questions are difficult to follow and comprehend. Others feel that it takes more time than ordinary questions. An example of such a question is one that reads, 'If one has answered question number 6 with 'yes' then skip question numbers 7 and 8 and answer questions 9 and 10. If one has answered question number 6 with a 'no', answer question numbers 7 and 8 and skip 9 and 10. It is equally unreasonable to ask the respondent to 'Rate from number 1 to 10 in order of preference of the following factors'.

Q25. Explain the Significance of Questionnaire.

Ans

The significance of questionnaire is as follows :

1. Quick and Easy to Create and Interpret

Questionnaires are relatively quick and easy to create, code and interpret (especially if closed questions are used). In addition, the respondent - not the researcher - does the time-consuming part of completing the questionnaire. A questionnaire is easy to standardise. For example, every respondent is asked the same question in the same way. The researcher, therefore, can be sure that everyone in the sample answers exactly the same questions, which makes this a very reliable method of research.

2. Accessibility

The researcher is able to contact large numbers of people quickly, easily and efficiently using a postal questionnaire (since all he/she has to do is identify the group that will be targeted and post them the list of questions). The postal or online questionnaire allows collecting data from a geographically dispersed sample group at a much lower cost than interviewing a similar sample. In addition, as not needed to be present to ask questions yourself, using questionnaires allows a larger sample to be investigated.

3. Potential Reduction in Bias

With a well-designed questionnaire there is little opportunity to introduce bias into the results as may be the cases with interviews, e.g., through the way respond to an answer, or your body language, or simply presence in observational studies. Researcher should be aware, however, that badly designed questionnaires can lead to bias in your data, and hence using a questionnaire does not automatically mean a reduction in bias.

4. Anonymity

Questionnaires can be used to explore potentially embarrassing areas more easily than other methods. The questionnaire can, e.g., be both anonymous and completed in privacy. This increases the chances of people answering questions honestly because they are not intimidated by the presence of a researcher. The presence of the researcher interested in certain sensitive issues, e.g., player violence, or the use of drugs, or cheating in sport, may inhibit the respondent. A postal questionnaire allows anonymity and may, therefore, improve the validity of your responses in certain cases.

5. Structured Data

Questionnaires tend to provide highly structured quantitative data that is easily comparable, either between subject group or between the same group studied over an extended time period. Such data is generally straightforward to convert into tables and charts, and to analyse statistically.

6. Increased Time for Respondents

Respondent-completed questionnaires allow the respondent to fill in the questionnaire at a convenient time if necessary, or to be able to go back to the questionnaire at a later time if they recall anything further.

Q26. What are the Factors considered while Design of Questionnaire ?

Ans :

Designing a questionnaire is not a simple job as it looks at first sight. A marketing researcher intending to collect primary data has to be extremely careful in deciding what information is to be collected. How many questions are to be formulate, what should be their sequence, what should be the working of each question, and what should be the layout of the questionnaire.

All these aspects need considerable time and effort of the marketing researcher. If he is able to develop a questionnaire suitable for his field investigation, he will find that his task of collecting the data has become much easier than otherwise.

1. Type of information to be collected

While attempting to design a questionnaire, the marketing researcher has to first as himself what type of information he needs from the survey. He should seriously consider this question as it will have considerable repercussion on the usefulness of the survey.

Generally there are different type of information in marketing research. The information could be one of more following types

1) Facts
2) Quasifacts
3) Awareness penetration of information
4) Opinions
5) Attitudes

About the information of questions should be collected for good questionnaire.

2. Types of questions

The second important aspect in the designing of a questionnaire is to be decided which type of decision are to be used. Question can be classified various types.

1) Open ended questions
2) Dichotomous questions
3) Multiple choice questions

A open ended or simply 'open] or 'free answer' question gives the respondent complete freedom to decide the form, length and detail of the answer. Open questions are prepared when the researcher is interested in knowing what is upper most in the mind of the respondent.

The dichotomous question has only tow answers in the form 'yes' or 'No' or false use or do not sue etc. an example of a dichotomous questions is

Do you use tobacco in any way ?

Yes - No -

There cannot be a third answer. However in some cases, there may be a third answer which may come from those who do not want to take a definite stand on way or the other.

For example, take the following question

Do you like to watch movies ?

Yes - No -

Neither like nor dis like -

The third alternative may be included so as to provide for the those respondents who do not have a positive preference or aversion to movies.

In the case of multiple choice questions, the respondent is offered tow or more choice for example the following is multiple choice question which of the following brand \ brands do you use for washing clothes?

Rin - , Det -, 501 Blur Bar -, Super 777 Bar

The respondent is likely to take more to answer a multiple – choice question as compared to dichotomous one. Also, more time is required in the editing, tabulation and interpretation of data.

3. Preparation of questionnaire

The next issue in preparation of a questionnaire is how to phrase the questions. The way in which a question is drafted is very important as a slightly suggestive wording would elicit a very different answer from the respondent. Consider, for example, the following question : don't you think that this is a sub standard product who do not have a definite opinion about the product, are likely to agree that it is of sub standard quality.

4. Order of questions

Another aspect that should receive the attention of the researcher is the sequence or order of questions to be contained in a questionnaire. Since, in the beginning, the researcher has to establish some rapport with the respondent, it is necessary that questions asked at the beginning are simple and thereby helpful in establishing the rapport. Difficult questions or those on sensitive issues should be relegated to the end of the questionnaire. Further questions of general type should be asked in the beginning while those which one specialized, needing some in depth information from the respondents, should de left to the end.

5. How many questions to be asked

The researcher has also decide how many questions one to be asked. We may add that the number of questions is not so important as the actual length of the questionnaire. Too lengthy a questionnaire is disadvantage for the respondent. Their opinion and reaction will be very helpful to marketing researcher.

6. Layout of the questionnaire

Finally, the researcher or some one on his behalf has to decide about the layout of the questionnaire. This implies that the document should be set in such a way that it leaves a favourable impression in the mind of the respondent. It should be neatly printed and the individual pages should not have to many questions so as to appear crowded.

7. Mail questionnaire

So far the discussion was confined to the designed of questionnaire to be filled in by personal interviews. In fact, the type of questionnaire to be designed depends on the type of survey. Broadly, there one three types of survey, personal, mail and telephone.

As far as the telephone survey is concerned, it is not commonly used in India. As such personal interview and mail survey are the only two methods. Since a mail survey need a questionnaire which should have some additional characteristics, it is necessary for marketing researcher.

8. Pre testing the questionnaire

Once the questionnaire is ready, it should be pre-tested of the questionnaire implies that it is tried out on a few respondents and their reaction to the questionnaire is observed. It helps the researcher decide whether any changes in question. Content or the wording of the questions are called for. If so, specific changes that are desirable can also be ascertained and incorporated in the questionnaire.

2.9 Sampling

Q27. What do you understand by sampling?

Ans :

Sampling may be defined as the selection of some part of an aggregate or totality on the basis of which a judgement or inference about the aggregate or totality is made. In other words, it is the process of obtaining information about an entire population by examining only a part of it.

In most of the research work and surveys, the usual approach happens to be to 'make generalizations' or to 'draw inferences' based on samples about the parameters of population from which the samples are taken. The researcher quite often selects only a few items from the universe for his study purpose. All this is done on the assumption that the sample data will enable him to estimate the population parameters.

The items so selected constitute what is technically called a sample, their selection process is called sample design and the survey conducted on the basis of sample is described a sample survey. Sample should be truly representative of population characteristics without any bias so that it may result in valid and reliable conclusions.

Q28. Explain the importance of sampling ?

Ans :

The importance of sampling is that you can determine the adequate respondents from the total number of target population. Thus, it will be used in the research study which should be adequate to warrant generalization of the findings to the target population. And the sample size represents the characteristics of the whole population (represen-tativeness of the sample). The advantages of sampling are: it is economical and practical; faster and cheaper; it can yield more comprehensive information; it is more accurate; and because of savings it permits in time and money, the sample survey makes possible the use of much larger and much more varied populations than would be possible for the same expenditure if one were making a complete enumeration.

Q29. What are the advantages of sampling ?

Ans :

Sampling ensures convenience, collection of intensive and exhaustive data, suitability in limited resources and better rapport. In addition to this, sampling has the following advantages also.

1. Low cost of sampling

If data were to be collected for the entire population, the cost will be quite high. A sample is a small proportion of a population. So, the cost will be lower if data is collected for a sample of population which is a big advantage.

2. Less time consuming in sampling

Use of sampling takes less time also. It consumes less time than census technique. Tabulation, analysis etc., take much less time in the case of a sample than in the case of a population.

3. Scope of sampling is high

The investigator is concerned with the generalization of data. To study a whole population in order to arrive at generalizations would be impractical.

Some populations are so large that their characteristics could not be measured. Before the measurement has been completed, the population would have changed. But the process of sampling makes it possible to arrive at generalizations by studying the variables within a relatively small proportion of the population.

4. Accuracy of data is high

Having drawn a sample and computed the desired descriptive statistics, it is possible to determine the stability of the obtained sample value. A sample represents the population from which its is drawn. It permits a high degree of accuracy due to a limited area of operations. Moreover, careful execution of field work is possible. Ultimately, the results of sampling studies turn out to be sufficiently accurate.

5. Organization of convenience

Organizational problems involved in sampling are very few. Since sample is of a small size, vast facilities are not required. Sampling is therefore economical in respect of resources. Study of samples involves less space and equipment.

6. Intensive and exhaustive data

In sample studies, measurements or observations are made of a limited number. So, intensive and exhaustive data are collected.

7. Suitable in limited resources

The resources available within an organization may be limited. Studying the entire universe is not viable. The population can be satisfactorily covered through sampling. Where limited resources exist, use of sampling is an appropriate strategy while conducting marketing research.

8. Better rapport

An effective research study requires a good rapport between the researcher and the respondents. When the population of the study is large, the problem of rapport arises. But manageable samples permit the researcher to establish adequate rapport with the respondents.

Q30. What are the stages of sampling Process ?

Ans :

The sampling process consists of five steps which are interrelated and are relevant in all the marketing research project aspects i.e., right from defining to presenting the outcomes.

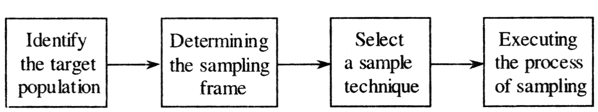


Fig. : Process of Sampling

1. Identifying the Target Population

This is the first step of sampling process. It involves collecting the element, which possesses the information, collected by the researcher, regarding which inferences are to be made.

The target population studies can be well defined in terms of units of sampling, time and extent. The elements can be an object from which the information can be obtained.

Example : Lakme wanted to assess the response of the customers, about a new line of compact powder and would like to take a sample of people who are above 19 years of age. In this case, it would be easier to collect a sample and the sampling unit would be the same as an element.

2. Sampling Frame

The sampling frame represents the elements of target population. The sampling frame includes sets of direction for the identification of the target population.

Example : A mailing list, purchased from an organisation, telephone book etc. Sometimes, it is possible that a list of population elements can be obtained but it might not include some elements of the population, it would lead to sampling frame error. The researcher is required to identify and correct the sampling error.

3. Selecting a Sampling Technique

It includes decisions of broader nature. There are several approaches available to the researcher such as the Bayesian approach, probability sampling or Non-Probability sampling approach.

In case Bayesian approach, the elements are selected in a sequence and are added to the sample. It also involves the process of collecting data, computation of statistics of sample and determination of its related costs. But this approach is not used widely as it does not provide complete information about elements of population and its related costs.

4. Sample Size

In this step, the size of the sample is determined, which is difficult and takes into consideration both the quantitative and qualitative techniques.

5. Executing the Processes of Sampling

This step involves a detailed specification of decisions related to the target population, sampling frame, techniques of sampling and sample size that are required to be implemented. This step requires the detailed information for all sampling design decisions.

Q31. Explain different types of Sampling Methods.

Ans : **(Imp.)**

Sampling Methods

Probability Sampling

- Random Sampling
- Systematic Sampling
- Stratified Sampling
- Multistage Sampling
- Purposive Sampling
- Cluster Sampling
- Multiple Sampling or Double Sampling

Non-Probability Sampling

- Incidental or Accidental Sampling
- Judgment Sampling
- Purposive Sampling
- Quota Sampling

A. Probability Sampling

G.C. Halmstadter, "A probability sample is one that has been used selected in such a way that every element chosen has a known probability of being included."

Probability sampling is of different types

1. Simple Random Sampling

It is one in which each element of the population has an equal and independent chance of being included in the sample i.e. a sample selected by randomization method is known as simple random sample and this technique is simple randomizing.

Randomization is done by using the following techniques:

(a) Tossing a coin

(b) Throwing a dice

(b) Lottery method

(d) Blind folded method

(c) Tippett's table method

Merits of Randomization

1. It requires the minimum knowledge of population.
2. It is free from subjectivity and free from personal error.
3. It provides appropriate data for one's purpose.
4. The observations of the sample can be used for inferential purpose.

Demerits of Randomization

1. It cannot ensure the representativeness of a sample.
2. It does not use the knowledge about the population.
3. Its inferential accuracy depends upon the size of the sample.

2. Systematic Sampling

Systematic sampling is an improvement over the simple random sampling. This method requires the complete information about the population. There should be a list of information of all the individuals of the population in any systematic way.

Now we decide the size of the sample:

Let the size of sample is = n and population size is = N

Now we select each N/n individual from the list and thus we have the desired size of sample which is known as systematic sample. Thus for this technique of sampling population should be arranged in any systematic way.

Merits

1. This is a simple method of selecting a sample.
2. It reduces the field cost.
3. Inferential statistics may be used.
4. Sample may be comprehensive and representative of population.
5. Observations of the sample may be used for drawing conclusions and generalizations.

Demerits

1. This is not free from error, since there is subjectivity due to different ways of systematic list by different individuals.
2. Knowledge of population is essential.
3. Information of each individual is essential.
4. This method can't ensure the representative-ness.
5. There is a risk in drawing conclusions from the observations of the sample.

3. Stratified Sampling

It is an improvement over the earlier methods. When we employ this technique, the researcher divides his population into strata on the basis of some characteristics and from each of these smaller homogenous groups (strata) draws at random a predetermined number of units. Researcher should choose that characteristic as criterion which seems to be more relevant in his research work.

Stratified sampling may be of three types;

(a) Disproportionate: Means that the size of the sample in each unit is not proportionate to the size of the unit but depends upon considerations involving personal judgement and convenience. This method of sampling is more effective for comparing strata which have different error possibilities. It is less efficient for determining population characteristics.

(b) Proportionate: It refers to the selection from each sampling unit of a sample that is proportionate to the size of the unit. Advantages of this procedure includes representativeness with respect to variables used as the basis of classifying categories and increased chances of being able to make comparisons between strata. Lack of information on proportion of the population in each category and faulty classification may be listed as disadvantages of this method.

(c) Optimum allocation: Stratified sampling is representative as well as comprehensive than other stratified samples. It refers to selecting units from each stratum. Each stratum should be in proportion to the corresponding stratum the population. Thus sample obtained is known as optimum allocation sample.

Merits

(i) It is a good representative of the population.

(ii) It is an improvement over the earlier technique of sampling.

(iii) It is an objective method of sampling.

(iv) Observations can be used for inferential purpose.

Demerits

(i) Serious disadvantage of this method is that it is difficult for the researcher to decide the relevant criterion for stratification.

(ii) Only one criterion can be used for stratification, but generally it seems more than one criterion relevant for stratification.

(iii) It is costly and time consuming method.

(iv) Selected samples may be representative with reference to the used criterion but not for the other.

(v) There is a risk of generalization.

4. Multiple or Double Repetitive Sampling

Generally this is not a new method but only a new application of the samplings. This is most frequently used for establishing the reliability of a sample. When employing a mailed questionnaire, double sampling is sometimes used to obtain a more representative sample. This is done because some randomly selected subjects who are sent questionnaires may not return them.

Obviously, the missing data will bias the result of the study, if the people who fail to reply the query differ in some fundamental way from the others in respect to the phenomenon being studied.

To eliminate this bias, a selected sample may be drawn at random from the non-respondents and the people interviewed to obtain the desired information. Thus this technique is also known as repeated or multiple sampling.

This double sampling technique enables one to check on the reliability of the information obtained from first sample. Thus, double sampling, where in one sample is analyzed and information obtained is used to draw the next sample to examine the problem further.

Merits

(i) Thus sampling procedure leads to the inferences of free determine precision based on a number of observations.

(ii) This technique of sampling reduces the error.

(iii) This method maintains the procedure of the finding evaluate the reliability of the sample.

Demerits

(i) This technique of sampling cannot be used for a large sample . It is applicable only for small sample.

(ii) This technique is time consuming and costly.

(iii) Its planning and administration is more complicated.

5. Multi Stage Sampling

This sample is more comprehensive and representative of the population. In this type of sampling primary sample units are inclusive groups and secondary units are sub-groups within these ultimate units to be selected which belong to one and only one group.

Stages of a population are usually available within a group or population, whenever stratification is done by the researcher. The individuals are selected from different stages for constituting the multi stage sampling.

Merits

(i) It is a good representative of the population.

(ii) Multistage sampling is an improvement over the earlier methods.

(iii) It is an objective procedure of sampling.

(iv) The observations from multi stage sample may be used for inferential purpose.

Demerits

(i) It is a difficult and complex method of sampling.

(ii) It involves errors when we consider the primary stages.

(iii) It is again a subjective technique of sampling.

6. Cluster Sampling

To select the intact group as a whole is known as a cluster sampling. In cluster sampling the sample units contain groups of element (cluster) instead of individual members or items in the population. Rather than listing all elementary school children in a given city and randomly selecting 15 % of these students for the sample, a researcher lists all of the elementary schools in the city, selects at random 15 % of these clusters of units, and uses all of the children in the selected schools as the sample.

Merits

(i) It may be a good representative of the population.

(ii) It is an easy method.

(iii) It is an economical method.

(iv) It is practicable and highly applicable in education.

(v) Observations can be used for inferential purpose.

Demerits

(i) Cluster sampling is not free from errors.

(ii) It is not comprehensive.

B. Non-Probability Sampling Method

Samples which are selected through non-random methods are called non probability samples. Depending upon the technique used it may be;

1. Incidental or Accidental Sampling

The term incidental or accidental applied to those samples that are taken because they are most frequently available i.e. this refers to the groups which are used as samples of a population because they are readily available or because the researcher is unable to employ more acceptable sampling methods.

Merits

(i) It is very easy method of sampling.

(ii) It is frequently used method in behavioural sciences.

(iii) It reduces the time, money and energy i.e. it is an economical method.

Demerits

(i) It is not representative of the population.

(ii) It is not free from errors.

(iii) Parametric statistics cannot be used.

2. Judgment Sampling:

This involves the selection of a group from the population on the basis of available information assuming as if they are representative of the entire population. Here group may also be selected on the basis of intuition or on the basis of the criterion deemed to be self-evident. Generally investigator should take the judgment sample so this sampling is highly risky.

Merits

(i) Knowledge of investigator can be best used in this technique of sampling.

(ii) This method of sampling is economical.

Demerits

(i) This technique is objective.

(ii) It is not free from errors.

(iii) It includes uncontrolled variation.

(iv) Inferential statistics cannot be used for the observation of this sampling, so generalization is not possible.

3. Purposive Sampling

The purposive sampling is selected by some arbitrary method because it is known to be representative of the total population, or it is known that it will produce well matched groups. The idea is to pick out the sample in relation to criterion which are considered important for the particular study. This method is appropriate when the study places special emphasis upon the control of certain specific variables.

Merits

(i) Use the best available knowledge concerning the sample subjects.

(ii) Better control of significant variables.

(iii) Sample groups data can be easily matched.

(iv) Homogeneity of subjects used in the sample.

Demerits

(i) Reliability of the criterion is questionable.

(ii) Knowledge of population is essential.

(iii) Errors in classifying sampling subjects.

(iv) Inability to utilize the inferential parametric statistics.

(v) Inability to make generalization concerning total population.

4. Quota Sampling

This combines both judgment sampling and probability sampling: on the basis of judgment or assumption or the previous knowledge, the proportion of population falling into each category is decided. Thereafter a quota of cases to be drawn is fixed and the observer is allowed to sample as he likes. Quota sampling is very arbitrary and likely to figure in municipal surveys.

Merits

(i) It is an improvement over the judgment sampling.

(ii) It is an easy sampling technique.

(iii) It is not frequently used in social surveys.

Demerits

(i) It is not a representative sample.

(ii) It is not free from errors.

(iii) It has the influence of regional , geographical and social factors.

5. Snowball Sampling

The term; snow ball sampling' has been used to describe a sampling procedure in which the sample goes on becoming bigger and bigger as the observation or study proceeds. The term snowball stems from the analogy of a snowball sample which would allow computation of estimates of sampling error and use of statistical test of significance.

For example, an opinion survey is to be conducted on smokers of a particular brand of cigarette. At the first stage, we may pick up a few people who are known to us or can be identified to be the smokers of that brand. At the time of interviewing them, we may obtain the names of other persons known to the first stage subjects. Thus the subjects go on serving an informant for the identification of more subjects and the sample goes on increasing.

Merit: Snowball sampling which is generally considered to be non-probabilistic can be converted into probabilistic by selecting subjects randomly within each stage.

Demerits: Sampling errors may creep in.

6. Purposive or Expert Choice Sampling

Samples are sometimes expressly chosen because, in the light of available information, these mirror some larger group with reference to one or more given characteristics. The controls in such samples are usually identified as representative areas (city, country, state, district), representative characteristics of individuals (age, sex, marital status, socio-economic status, race) or types of groups (administrator, counselors, teachers etc.).

These controls may be further sub-divided by specified categories within classes such as amount of training, years of experience or attitudes towards a specific phenomenon. Up-to this stage, these controls are somewhat similar to those used in satisfaction. Purposive sampling differs from stratified random sampling in that the actual selection of the units to be included in the sample in each group is done purposively rather than by random method.

Q32. What are the differences between Probability and Non-Probability Sampling

Ans :

	Probability Sampling	Non-probability Sampling
1.	It is a method of sampling which gives the probability that a sample is representative of population.	In the absence of any idea of probability the method of sampling is known as non-probability sampling.
2.	Probability sampling is generally used in fundamental research in which the purpose is to generalize the results.	It is generally used in action researches in which one studies a class without any generalization purpose.
3.	It refers from the sample as well as the population.	There is no idea of population.
4.	Every individual of the population has equal probability to be taken into the sample.	There is no probability of selecting any individual.
5.	It may be representative of the population.	It has free distribution.
6.	Its observations (data) are used for the inferential purpose.	The observations are not used for generalization purpose.
7.	Inferential or parametric statistics are used.	Non-inferential or non-parametric statistics are used.
8.	There is a risk of drawing conclusion.	There is no risk for drawing conclusions.
9.	It is based on Law of probability sampling i.e. Law of Statistical Regularity and Law of Inertia of the Large Sample.	It is not based on law of probability sampling.

2.9.1 Applications of Sampling

Q33. What are the Applications of Sampling?

Ans : **(Imp.)**

1. Reduced Cost of Enquiry

Sampling usually results in reduction of cost in terms of money and in terms of man-hours. Although the amount of labor and expenses involved in collecting information are generally greater per unit of sample than the total cost of the sample, survey is expected to be much lower than that of census. Since in most of the cases our resource are limited in terms of money and the time within which the results of the survey should' be obtained, it is usually desirable to resort to sampling rather than complete enumeration.

2. Saving in Time and Labour

Since only a part of the population is to be inspected and examined, the sample method results in considerable amount of saving in time and labour. There is saving in time not only in conducting the sampling enquiry but also in the processing, editing and analysing the data. This is a very sensitive and important point for a statistical investigation where the results are urgently and quickly needed.

3. Sometimes the Only Method Possible

When the investigation entails destruction of material, example the strength of a bullet, the life of a lamp, sampling is the only practical way of assessing the quality of the whole lot. Also if the universe is too large or infinite or spread over a large geographical area, it is difficult to collect information about each unit and there is no alternative but to resort to sampling. For example, fish in a river or number of wild animals in a dense forest can be studied by sample enumeration only.

4. Administrative Convenience

A complete census required a very huge administrative setup involving lot of personnel, trained investigators and above all the coordination between the various operating agencies. On the other hand, the organization and administration of a sample survey is relatively much convenient as it requires less staff and the field of enquiry is also limited.

5. Detailed Inquiry

With small data, it is possible to have greater precision and an in-depth study, because a more detailed information can be sought from a small group of respondents.

6. Results More Reliable

Conclusions and results obtained by sampling enumeration are generally more reliable than those obtained by census enumeration because,

(a) With a small data to process, there are fewer chances of non-sampling statistical errors. The sampling errors would, of course, be there but it is possible to estimate and control them.

(b) Highly trained personnel and specialized equipment can be employed for scientific processing and analysis of relatively limited data. As such more sophisticated statistical techniques can be used to obtain more accurate and reliable results. These can be employed in census enumeration only with the aid of computer.

7. More Scientific

The size method of sampling is scientific and is not based on expediency or more tradition. This method has full justification for the expenditure involved. Also it is easy to guard against incomplete and inaccurate returns in relatively limited enumeration. There can be easy follow-up in case of non-response or incomplete response.

8. Hypothetical Population

In case of hypothetical population, as for example in the problem of throwing a die or tossing a coin where the process may continue a large number of times or infinitely, the sampling procedure is the only scientific technique of estimating the parameters of the population.

Short Question and Answers

1. Components of Research Design.

Ans :

i) **Title of the Study:** Should be brief, precise, and project the scope in generalized terms.

ii) **Statement of the Problem:** Should be unambiguous, precise, and the usage should be clear, simple, and concise.

iii) **Review of Previous Studies:** Should be a brief survey of a relevant literature in the subject concerned and presented subjectwise and reviewed critically and pinpoint the stage from where further research is called for.

iv) **Definition of Concept or Theoretical Principles involved if any:** Should be defined in general terms and be linked with the study.

v) **Coverage and Scope of the Study:** Should consider, geographical, temporal, and functional dimensions.

vi) **Objectives of the Study:** Should explain the main purpose precisely and may be in the form of questions or an explanation to a particular issue.

2. What are the Requirements for a Good Research Design.

Ans :

i) **Well Defined Problems:** Nature and scope of the problem to be studied must be stated clearly, or say must be well defined and formulated.

ii) **Clarity in Formulation:** If any hypothesis is to be tested it must be clearly formulated.

iii) **Testable:** The research design must adequately answer the research questions. And test the hypothesis.

iv) **Identifying Variables:** Relevant variables must be clearly identified and operationalized. Adequate methods of collecting the information and methods of logically deriving the conclusion must be developed. Only then control of variance is possible.

v) **Serve the Validity Needs:** The research design must be structured in a manner that it fulfils the need of internal and external validity.

3. Cross-sectional

Ans :

A Cross-Sectional study is concerned with a sample of elements from a given population. Thus, it may deal with households, dealers, retail stores, or other entities. Data on a number of characteristics from the sample elements are collected and analyzed.

Cross-sectional studies are of two types: Field Studies and Surveys. Although the distinction between them is not clear, there are some practical differences which need different techniques and skills.

(a) Field Studies

Field studies are *ex-post-facto* scientific enquires that aim at finding the relations and interrelations among variables in a real setting. Such studies are done in life situations like communities, schools, factories, organizations, and institutions.

Field studies have their strengths and weaknesses. One major strength is that they are close to real life, and they cannot be criticized on the ground that they are remote from real settings or are artificial. Further, in real settings, variables exert their influence fully and, as such, the strength of variables is another advantage of field studies. Field studies are also strong in their heuristic quality.

Field studies are also subject to certain weaknesses. Such studies are scientifically inferior to laboratory and field experiments. One of their major weaknesses is their *ex-post- facto* character. As a result, interrelations among variables are weaker than they are in laboratory experiments. Another weakness is the lack of precision in the measurement of variables.

(b) Survey Research

Another type of cross-sectional study is Survey Research. A major strength of Survey Research is its wide scope. Detailed information can be obtained from a sample of a large population. Besides, it is economical as more information can be collected per unit of cost. Also, it is obvious that a sample survey needs less time than a census enquiry.

Despite these advantages of survey research, it is subject to certain limitations. Generally, survey research does not penetrate below the surface as more emphasis is given to the extent of information sought rather than to an in-depth analysis. Another disadvantage is that, Survey Research demands more time and more money, specially when it is conducted on a arge scale. Another limitation of survey research is that the interview may makes the respondent alert and cautious and he may not answer the questions in a natural manner. Such answers will make the survey invalid.

4. Advantages of Longitudinal Studies

Ans :

There are several advantages of using panel data,

1. Such data will enable the researcher to undertake detailed analysis.
2. Another advantage of the panel is that more comprehensive data could be obtained as individuals or families included in the panel are those who have accepted to provide data periodically. As panel members are willing persons, more data can be collected.
3. The other advantage is that panel data has been found to be more accurate than data collected through survey.
4. Costs of data collection through panels are generally lower than collected through personal interviews.

5. Advanteges of Primary Data

Ans :

1. Reliability

The information collected for primary data is more reliable than those collected from the secondary data because this information is collected directly from the respondents.

2. Availability of a Wide Range of Techniques

There are lot of techniques that can be employed, which means that all information necessary can be obtained by using the appropriate techniques, enabling all areas of the research topic to be answered and investigating thoroughly and effectively.

3. Addresses Specific Research Issues

The organisation asking for the research has the complete control on the process and the research is streamlines as far as its objectives and scope is concerned. Researching company can be asked to concentrate their efforts to find data regarding specific market rather than concentration on mass market. Primary research is designed to collect the information the marketer wants to know, and report it in ways that benefit the marketer.

4. Greater Control

Not only does primary research enable the marketer to focus on specific issues, it also enables the marketer to have a higher level of control over how the information is collected. In this way the marketer can decide on such issues as size of project (e.g., how many responses), location of research (e.g., geographic area) and timeframe for completing the project.

5. Efficient Spending for Information

Unlike secondary research, where the marketer may spend for information that is not needed, primary data collection focuses on issues specific to the researcher and improves the chances that research funds will be spent efficiently.

6. Disadvantages of Secondary Data

Ans :

It is however difficult to find secondary data that exactly fits into the needs of some specific research investigation. The problems experienced in respect of secondary sources can be in terms of :

1. Relevance

The data may not fit into the needs of investigation. There may be difference in the units of measurement; there may be surrogated data; discrepancy of classes and data may pertain to some other period of time.

2. Accuracy

It is observed that it is rather difficult to measure the degree of approximations used in the collection of information as well as the competence of the investigator in motivating the persons to supply the desired information.

3. Existence of Obsolete Information

Information may be outdated or obsolete.

4. Non-Disclosure of Research Findings

All the findings of a research study may not be made public.

5. Seldom Catering to the Need

The available data may not suit the current purpose of research, due to incompleteness, generalities and so on.

6. Other Limitations

i) There may be difficulties in the identification of the source.

ii) Errors may be there in recording or transferring information from secondary sources.

iii) The facilities or capabilities of the agency that originally collect the data might be questionable.

7. Define Questionnaire.

Ans :

Questionnaire is a data collection instrument. A questionnaire is a prepared set of questions (or measures) used by respondents or interviewers to record answers (data). The term questionnaire usually refers to a self- administered process where by the respondent himself reads the question and records his answers without the assistance of an interviewer. This is a narrow definition of a questionnaire.

A questionnaire is a method of obtaining specific information about a defined problem so that the data, after analysis and interpretation, results in a better appreciation of the problem. A questionnaire form, which has to be completed by an interviewer, is often referred as schedule.

The success of collecting data either through the questionnaire method or through the schedule method depends largely on the proper design of the questionnaire. This is a specialised job and requires high degree of skill, experience, through knowledge of the research topic, ability to frame questions and a great deal of patience. There are no hard and fast rules in designing the questionnaire.

Questionnaire design is only one phase of several interrelated business research steps. But it is a very important phase because data collected with questionnaires is used to improve decision-making. Moreover, it is a structured framework consisting of a set of questions and scales designed to generate primary data.

8. Structured, Non-Disguised Questionnaire.

Ans :

Most questionnaire studies made in marketing research are of the first type- they are structured and are not disguised. If the sales manager for a musical instrument company wants to find out how many and what type of people play various types of instruments, a formal list of questions may be set up that asks directly about the ownership and playing of various instruments. Each of a selected group of persons is then asked this set of questions in the given sequence. Answers are frequently limited to a list of alternatives, which is stated or implied. Several questions taken from an actual survey of this type are given below.

Structured, non-disguised studies can be handled by telephone, mail, or personal interview. They are subject to the three limitations of the questionnaire -method-respondents may be unable to furnish the information desired, they may be unwilling to furnish it, or the questioning process may tend to stimulate incorrect or misleading answers.

9. Sampling

Ans :

Sampling may be defined as the selection of some part of an aggregate or totality on the basis of which a judgement or inference about the aggregate or totality is made. In other words, it is the process of obtaining information about an entire population by examining only a part of it.

In most of the research work and surveys, the usual approach happens to be to 'make generalizations' or to 'draw inferences' based on samples about the parameters of population from which the samples are taken. The researcher quite often selects only a few items from the universe for his study purpose. All this is done on the assumption that the sample data will enable him to estimate the population parameters.

The items so selected constitute what is technically called a sample, their selection process is called sample design and the survey conducted on the basis of sample is described a sample survey. Sample should be truly representative of population characteristics without any bias so that it may result in valid and reliable conclusions.

10. What are the differences between Probability and Non-Probability Sampling

Ans :

S.No.	Probability Sampling	Non-probability Sampling
1.	It is a method of sampling which gives the probability that a sample is representative of population.	In the absence of any idea of probability the method of sampling is known as non-probability sampling.
2.	Probability sampling is generally used in fundamental research in which the purpose is to generalize the results.	It is generally used in action researches in which one studies a class without any generalization purpose.
3.	It refers from the sample as well as the population.	There is no idea of population.
4.	Every individual of the population has equal probability to be taken into the sample.	There is no probability of selecting any individual.
5.	It may be representative of the population.	It has free distribution.
6.	Its observations (data) are used for the inferential purpose.	The observations are not used for generalization purpose.
7.	Inferential or parametric statistics are used.	Non-inferential or non-parametric statistics are used.
8.	There is a risk of drawing conclusion.	There is no risk for drawing conclusions.
9.	It is based on Law of probability sampling i.e. Law of Statistical Regularity and Law of Inertia of the Large Sample.	It is not based on law of probability sampling.

UNIT III

(a) **Tabulation** of Univariate, Bivariate and multivariate data, Data classification and tabulation, Diagrammatic and graphical representation of data. One dimensional, two dimensional and three-dimensional diagrams and graphs

(b) **Small Sample Tests** - t-Distribution-properties and applications, testing for one and two means, paired t-test.

3.1 Tabulation of Univariate, Bivariate and Multivariate Data

Q1. What is tabulation ? Explain the significance of tabulation.

Ans : **(Jan.-20, Imp.)**

One of the simplest and most revealing devices for summarizing data and presenting them in a meaningful fashion is the statistical table . A table is a systematic arrangement of statistical data in columns and rows. Rows are horizontal arrangements whereas columns are vertical ones. The purpose of a table is to simplify the presentation and to facilitate 'comparisons. The simplification results from the clear-cut and systematic arrangement, which enables the reader to quickly locate desired information.

The significance of tabulation will be clear from the following points:

(i) It simplifies complex data

When data are tabulated all unnecessary retails and repetitions are avoided. Data are presented systematically in columns and rows. Hence, the reader gets a very clear idea of what the table presents. There is thus a considerable saving in time taken in understanding what is represented by the data and all confusion is avoided. Also a large amount of space is saved because of non-duplicating of his headings and designations; the description at the top of a column serves for all the terms beneath it.

(ii) It facilitates comparison

Tabulation facilitates comparison. Since a cable is divided into various parts and for each part there are totals and sub-totals, the relationship between different parts of data can be studied much more easily with the help of a table than without it.

iii) It gives identity to the data

When the data are arranged in a table with a title and number they can be distinctly identified and can be used as a source reference in the interpretation of a problem.

iv) It reveals patterns

Tabulation reveals patterns within the figures which cannot be seen in the narrative form. It also facilitates the summation of the figures if the reader desires to check the totals.

Q2. Explain different parts of a table.

Ans : **(Jan.-20, Imp.)**

The number of parts of a table varies from case to case depending upon the given data. However, the main parts of a table in general are discussed here.

1. Table Number

Each table should be numbered. There are different practices with regard to the place where this number is to be given. The number may be given either in the centre at the top above the title or inside of the title at the top or in the bottom of the table on the left-hand side. However if space permits the table number should be given in the, centre as is shown in the specimen table given on page. Where there, are many columns, it is also desirable to number each column so that easy reference to it is possible.

2. Title of the Table

Every table must be given suitable title. The title is a description of the contents of the table. A complete title has to answer the questions what, where and when in that sequence. In other words,

(a) what precisely are the data in the table (i.e., what categories of statistical data are shown)?

(b) where the data occurred (i.e., the precise geographical, political or physical area covered)?

(c) when the data occurred (i.e., the specific time or period covered by the statistical materials in the table) ?

The title should be clear, brief and self-explanatory. However, clarity should not be sacrificed for the sake of brevity. Long titles cannot be read as promptly as short titles, but at times they may have to be used for the sake of clarity. The title should be so worded that it permits one and only one interpretation. It should be in the form of a series of phrases rather than complete sentences. Its lettering should be the most prominent of any lettering on the table.

3. Caption

Caption refers to the column headings. It explains what the column represents. It may consist of one or more column headings. Under a column heading there may be sub-heads. The caption should be clearly defined and placed at the middle of the column. If the different columns are expressed in different units, the units should be mentioned with the captions. As compared with the main part of the table the caption should be shown in smaller letters. This helps in saving space.

4. Stub

As distinguished from caption, stubs are the designations of the rows or row headings. They are at 'the extreme left and perform the same function for the horizontal rows of numbers in the table as the column headings do for the vertical columns of numbers. The stubs are usually wide than column headings but should be kept as narrow as possible without sacrificing precision and clarity of statements.

5. Body

The body of the table contains the numerical information. This is the most vital part of the table. Data presented in the body arranged according to description are classifications of the captions and stubs.

6. Headnote

It is a brief explanatory statement applying to all or a major part of the material in the table, and is placed below the point centered and enclosed in brackets. It is used to explain certain points relating to the whole table that have not been included in the title nor in the captions or stubs. For example, the unit of measurement is frequently written as a headnote, such as "in thousands" or "in million tonnes" or "in crores", etc.

7. Footnotes

Anything in a table which the reader may find difficult to understand from the title, captions and stubs should be explained in footnotes. If footnotes are needed they are placed directly below the body of the table. Footnotes are used for the following main purposes:

(a) To point out any exceptions as to the basis of arriving at the data, for example, sales recorded at 'ex-factory price' for some of the entries and at 'delivered price' for others. Any heterogeneity in the data recorded must be disclosed to avoid wrong conclusions.

(b) Any special circumstances affecting the data, for example, strike, lock-out, fire, etc.

(c) To clarify anything in the table.

(d) To give the source in case of secondary data. The reference to the source should be complete in itself. For example, if the data is obtained from some periodical, its name, date of publication, page number, table number, etc., should be mentioned so that if the user wishes to check the data from the original source or considers later data from the same source he will know where to look for the information.

Format of a Table

Title	Headnote
Stub Heading	Caption heading – Column heading
Stub Entries	Body
Footnotes	Table Number

Q3. Explain the general rules for tabulation.

Ans :

- The table should suit the size of the paper usually with more rows than columns. In making a suitable layout it may be necessary to alter the original design. The alteration often consists in changing the rows to columns or the other way round. For this reason, it is desirable to make a rough draft of the table before the figures are entered in it. Space must be allowed for reference or any other matter which is to be included in the table.
- In all tables the captions and stubs should be arranged in some systematic order. It would make the table easier to read and allow more important items to be emphasized. The arrangement of items basically depends upon the type of data. However, the principal bases for arranging items are the following:
  (a) **Alphabetical** i.e., arrangement according to alphabets. The type of arrangement is very common in general purpose or reference tables.
  (b) **Chronological** i.e., arrangement according to time. This type of arrangement is of particular value in presenting historical data.
  (c) **Geographical** i.e., arrangement of data in certain territorial units such as countries, cities, districts, etc.
  (d) **Conventional** or arrangement in a customary order such as men, women and children or Hindus, Muslims, Sikhs and Christians.
  (e) Items may be arranged according to size, i.e., the numerical importance of the items, the largest items being given first and the smallest in the last. This arrangement may be reversed, if necessary.
- The point of measurement should be clearly defined and given in the table such as income in rupees or weight in pounds, etc.
- Figures should be rounded off to avoid unnecessary details in the table and a footnote to this effect should be given. For example, the figures may be taken to the nearest rupees and paise be eliminated.
- If certain figures are to be emphasized they should be in distinctive type or in a 'box' or circle or between thick lines.
- The table should not be overloaded with details. If many characteristics are to be shown it is not necessary to load them all in one table; rather a number of tables should be prepared, each table complete in itself and serving a particular purpose.

- A column entitled 'miscellaneous' should be added for data which do not fit in the classification made.
- The arrangement of the table should be logical and items related to each other should be placed near about and, if possible, in the same group. Derivative figures such as totals, averages and percentages should be placed near the original figures. Columns and rows should be numbered for identification since reference is more easily given by quoting numbers than the title of the column.
- Percentage and ratios should be computed and shown, if necessary. Frequently, figures in tables become more meaningful if expressed as percentages or (less often) as ratios. In constructing a table, therefore, it is important to decide whether or not it can be improved in this way. If it can, additional column should be inserted in the table and the percentages (or ratios) computed and entered. Such percentages and ratios are sometimes called derived statistics.
- Where standard classifications have been prepared it is usually desirable to employ them, as they are superior to hastily constructed individual classification.
- Indicate a zero quantity by a zero, and do not use zero to indicate information which is not available. If it is not available, show this fact by the letter N.A. or dash (-).
- Abbreviations should be avoided especially in titles and headings. For example "yr" should not be used for 'year'.
- Be explicit. The expression "etc." is bad form in a table, since the reader may not readily discover what it refers to. In fact clarity is the most important feature of tabular presentation of any kind of statistical data.

Q4. Explain different types of Tables.

Ans :

Tables can be classified according to their purpose, stage of enquiry, nature of data or number of characteristics used. On the basis of the number of characteristics, tables may be classified as follows:

1. Simple or one-way table
2. Two way table
3. Manifold table

1. Simple or one-way Table

A simple or one-way table is the simplest table which contains data of one characteristic only. A simple table is easy to construct and simple to follow. For example, the blank table given below may be used to show the number of adults in different occupations in a locality.

The number of adults in different occupations in a locality.

Occupations	No. Of Adults
Total	

2. Two-way Table

A table, which contains data on two characteristics, is called a two way table. In such case, therefore, either stub or caption is divided into two co-ordinate parts. In the given table, as an example the caption may be further divided in respect of ‘ sex’ . This subdivision is shown in two-way table, which now contains two characteristics namely, occupation and sex.

The umber of adults in a locality in respect of occupation and sex.

Occupation	No. of Adults		Total
	Male	Female	
Total			

3. Manifold Table

Thus, more and more complex tables can be formed by including other characteristics. For example, we may further classify the caption subheadings in the above table in respect of “marital status”, “ religion” and “socio-economic status” etc. A table ,which has more than two characteristics of data is considered as a manifold table. For instance , table shown below shows three characteristics namely, occupation, sex and marital status.

Occupation	No. of Adults						Total
	Male			Female			
	M	U	Total	M	U	Total	
Total							

Foot note: M Stands for Married and U stands for unmarried.

Manifold tables, though complex are good in practice as these enable full information to be incorporated and facilitate analysis of all related facts. Still, as a normal practice, not more than four characteristics should be represented in one table to avoid confusion. Other related tables may be formed to show the remaining characteristics.

Q5. Explain univariate tabulation with an example.

Ans :

In statistics tabulations can be either univariate, bivariate and multivariate in nature. Uni-variate analysis is one of the simplest form of statistical analyses which includes only one question. This one question has to be tabulated and this is called as a univariate tabulation. Univariate tables are usually used when the question itself gives the vital important for the analysis. Univariate table can be explained as follows,

Example

Find the percentage of the customers opinions regarding the product performance.

Response Category	Number of Responses	Percentage
Excellent	40	26.66
Good	50	33.33
Fair	30	20
Poor	30	20
Total	**150**	**100**

Sol :

33.33% of the responses opinion is good regarding the product and 20% opined to be fair and 20% opined for poor performance of the product.

Q6. What do you understand bivariate / multivariate tabulation.

Ans : **(Nov.-21)**

Bivariate/Multivariate Tabulation

Even though majority of the researchers use univariate tables based on the responses to only one question yet, they always attempt to get more helpful information by counting the responses to two or more queries either together or individually.

Example

It might be possible that women are not smokers. But women from a specific high income group can be smokers. In order to analyze this, we need to club the category of 'women' and 'high income group'.

While on the other hand, the univariate tabulation leads in frequency distribution of responses to one question, the bivariate tabulation results in a frequency distribution of responses to two or more questions. The bivariate/multivariate tabulation comprises a table having two or more rows and two or more columns.

In order to include the responses to a third question. The rows and columns needs to be divided. Otherwise, a seperate bivariate tables can also be created based on the number of classes, in the third question. Depending on the class of responses to the third question, a separate bivariate table is prepared which includes the responses to the first two question.

An example of bivariate/multivariate tabulation that represents the first alternative is as follows,

M/F	Age	Agree	Disagree	Not Sure	Total
Men	Below 35	-	-	-	-
	Above 35	-	-	-	-
	Total	-	-	-	-
Women	Below 35	-	-	-	-
	Above 35	-	-	-	-
	Total	-	-	-	-

From the above table, it is being observed that the table can be considered as two separate bivariate tables one for women and another for men.

In bivariate/multivariate tabulations, the main problem is in choosing the correct combinations of questions out of the large number of available combinations. It is essential for a researcher to select the right combination clearly. Otherwise it would be a waste of time in creating unnecessary or complex tables which not only creates the chaos among the reader but also provides unnecessary information. The objectives of the research can help in selecting the appropriate combination.

3.2 Data Classification and Tabulation

Q7. Define classification. What are the objectives of classification?

Ans : **(Imp.)**

Classification is the grouping of related facts into classes. Facts in one class differ from those of classification. Sorting facts on one basis of classification and then on another basis is called cross-classification. This process can be repeated as many times as there are possible bases of classification. Classification of data is a function very similar to that of sorting letters in a post-office. It is well known that letters collected in a post-office are sorted into different lots on a geographical basis, i.e., in accordance with their destinations such as Mumbai, Calcutta, Kanpur, Jaipur, etc. They are then put into separate bags, each containing letters with a common characteristic, viz., having the same destination. Classification of statistical data is comparable to the sorting operation.

The principal objectives of classifying the data are :

- To condense the mass of data in such a manner that similarities and dissimilarities can be readily apprehended. Millions of figures can thus be arranged in a few classes having common features.
- To facilitate comparison.
- To pinpoint the most significant features of the data at a glance.
- To give prominence to the important information gathered while dropping out the unnecessary elements.
- To enable a statistical treatment of the material collected.

Q8. Explain different types of classification?

Ans : **(Imp.)**

Broadly , the data can be classified on the base of following four criteria:

(i) Geographical, i.e., area-wise, e.g., cities, districts, etc.

(ii) Chronological, i.e., on the basis of time.

(iii) Qualitative, i.e., according to some attributes.

(iv) Quantitative, i.e., in terms of magnitudes.

(i) Geographical Classification

In this type of classification data are classified on the basis of geographical or vocational differences between the various items, like States, cities, regions, zones, areas, etc. For instance, the production of foodgrains in India may be presented State-wise in the following manner:

State-wise Estimates of Production of Foodgrains

Name of Slate	Total Foodgrains (Thousand tonnes)
Andhra Pradesh	1093.7
Bihar	12899.0
Haryana	11334.7
Punjab	21148.9
Uttar Pradesh	41828.6
All India	1,92,433.6

Geographical classifications are usually listed in alphabetical order for easy reference. Items may also be listed by size to emphasize the important areas as in ranking the States by population. Normally, in reference table the first approach is followed and in summary tables the second approach is followed.

(ii) Chronological Classification

When data are observed over a period of time the type of classification is known as chronological classification. For example, we may present the figures of population (or production, sales, etc.) as follows:

Population of India From 1941 To 1991

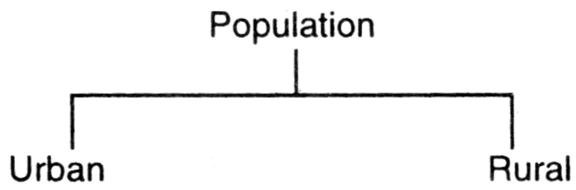
Year	Population (in crores)
1941	31.87
1971	54.82
1951	36.11
1981	68.33
1961	43.92
1991	84.63

Time series are usually listed in chronological order, normally starting with the earliest period. When the major emphasis falls on the most recent events, a reverse time order may be used.

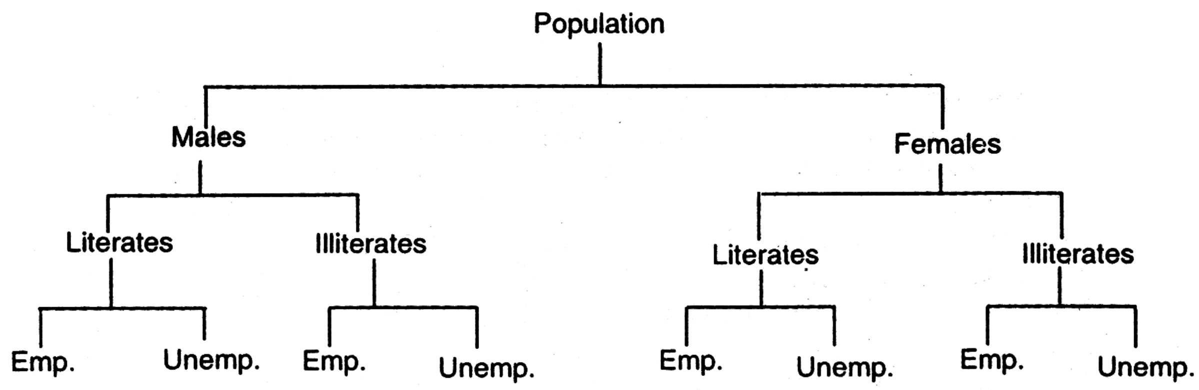
(iii) Qualitative Classification

In qualitative classification data are classified on the basis of some attribute or quality such as sex, colour of hair, literacy, religion, etc. The point to note in this type of classification is that the attribute under study cannot be measured: one can only find out whether it is present or absent in the units of the population under study.

For example, if the attribute under study is population, one can find out how many persons are living in urban area and how many in rural area. Thus when only one attribute is studied two classes are formed, one possessing the attribute and the other not possessing the attribute. This type of classification is known as simple classification. For example, the population under study may be divided into two categories as follows:



In a similar manner, we may classify population on the basis of sex, i.e., into males and females, or literacy, i.e., into literate and illiterate, and so on. The type of classification where only two classes are formed is also called two-fold or dichotomous classification. If instead of forming only two classes we further divide the data on the basis of some attribute or attributes so as to form several classes, the classification is known as manifold classification. For example, we may first divide the population into males and females on the basis of the attribute 'sex' ; each of these classes may be further subdivided into 'literate' and 'illiterate' on the basis of the attribute 'literacy'. Further classification can be made on the basis of some other attribute, say, employment. An example of manifold classification is given here:



Note. Emp. indicates Employed and Unemp. indicates Unemployed.

(iv) Quantitative Classification

Quantitative classification refers to the classification of data according to some characteristics that can be measured, such as height, weight, income, sales, profits, production, etc. For example, the students of a college may be classified according to weight as follows:

Weight (in lb.)	No. of Students
90 - 100	50
100 - 110	200
110 - 120	260
120 - 130	360
130 - 140	90
140 - 150	40
Total	**1,000**

Such a distribution is known as empirical frequency distribution or simple frequency distribution.

In this type of classification, there are two elements, namely (i) the variable, i.e., the weight in the above example, and (ii) the frequency, i.e., the number' of students in each class. There were 50 students having weight ranging from 90 to 100 lb., 200 students having weight ranging from 100 to 110 lb., and so on.

PROBLEMS ON TABULATION

1. Point out the mistakes in the following table drawn to show the distribution of population, according to sex, age and literacy.

Sex	0 to 25	25 to 50	50 to 75	75 to 100
Males				
Females				

Sol :

All the characteristics are not revealed in the above table; the characteristic of literacy has been completely ignored. Even otherwise the table needs to be re-arranged as follows :

Table Showing the Distribution of Population

According to Age, Sex and Literacy

Age groups	Literates			Illiterates			Total		
	M	F	Total	M	F	Total	M	F	Total
0 to 25	–	–	–	–	–	–	–	–	–
25 to 50	–	–	–	–	–	–	–	–	–
50 to 75	–	–	–	–	–	–	–	–	–
75 to 100	–	–	–	–	–	–	–	–	–
Total	–	–	–	–	–	–	–	–	–

2. Draft a blank table to show the distribution of personnel in a manufacturing concern according to :

(a) Sex : males and females

(b) Three grades of salary : below Rs. 5,000, Rs. 5000 – 10,000, Rs. 10,000 and above,

(c) Two periods : 1999 and 2000

(d) Three age groups : below 25, 25 and under 40,40 and over.

Sol :

Table Showing distribution of Personnel According to Sex

Salary and Age Groups for two years

Year	Age Groups	Below Rs. 5000			Salary Grade Rs. 5,000 1,0000			Rs. 10,000 and above		
		M	F	Total	M	F	Total	M	F	Total
1999	Below 25	–	–	–	–	–	–	–	–	–
	25 and under 40	–	–	–	–	–	–	–	–	–
	40 and above									
	Total	–	–	–	–	–	–	–	–	–
Zero	Below 25	–	–	–	–	–	–	–	–	–
	under 40	–	–	–	–	–	–	–	–	–
	40 and above									
	Total	–	–	–	–	–	–	–	–	–

3. Present the following information in a suitable tabular form :

In 1985 out of a total of 1, 750 workers of a factory, 1,200 were members of a trade union . The number of women employees was 200 of which 175 did not belong to a trade union. In 1990 the number of union workers increased to 1,580 of which 1, 290 were men. On the other hand the number of non-union workers fell down to 208 of which 180 were men.

In 1995 there were 1,800 employees who belonged to a trade union and 50 who did not belong to a trade union and 50 who did not belong to a trade union. Of all the employees in 1995, 300 were women of whom only 8 did not belong to a trade union.

Sol :

Table Showing Sex - wise Distribution of Union

and Non -Union Members For 1985, 1990 and 1995

	1985			1990			1995		
Category	M	F	Total	M	F	Total	M	F	Total
Members	1,175	25	1,200	1,290	290	1,580	1,508	292	1,800
Non - members	375	175	550	180	28	208	42	8	50
Total	1,550	200	1,750	1,470	318	1,788	1,550	300	1,850

Q9. What are the difference between Classification and Tabulation?

Ans :

Classification	Tabulation
It is the basis for tabulation	It is the basis for further analysis
It is the basis for simplification	It is the basis for presentation
Data is divided into groups and subgroups on the basis of similarities and dissimilarities.	Data is listed according to a logical sequence of related characteristics
data are separated and grouped based on a property of the data common to all values	Data is arranged into columns and rows based on characteristics /properties, or indicators.
emphasize on the presentation aspects of the data	used as a means of sorting of data for further analysis

Q10. What is the purpose of classifying statistical data ? Explain the different methods of classification.

Ans :

Purpose of Classification of Statistical Data

The classification of statistical data basically fulfil r the following purposes,

1. It summarizes the raw data and presents it in an appropriate form which is best suitable for statistic! analysis.
2. It shows the important characteristics of data and eliminates the difficulties.
3. It helps in making comparisons and conclusions out of the data.
4. It gives the information regarding the mutual relation of the elements of data set.
5. It segregates the various elements of data set by bringing out the similarities and dissimilarities among them.

This is useful in carrying out statistical analysis.

Methods of Classification

The three major methods of classification of data are as follows,

1. Exclusive method
2. Inclusive method
3. Open-End class interval method.

1. Exclusive Method

In exclusive method, the upper limit of one class would be the lower limit of the successive class when the class intervals in a frequency distribution are fixed. This is explained below,

Marks	Number of Students
5 - 10	6
10 - 15	9
15 - 20	12
20 - 25	20
25 - 30	16
30 - 35	28
35 - 40	13

2. Inclusive Method

In inclusive method, the upper limit of one class is also included in the similar class itself when the class intervals in a frequency distribution are fixed. This is explained below,

Weekly Wages (in `)	No. of Workers
51 - 100	5
101 - 150	14
151 - 200	25
201 - 250	28
251 - 300	26
301 - 350	24
351 - 400	32

3. Open-End Class Interval Method

An open-end class interval method does not provide any upper limit and lower limit. This is explained below,

Monthly Income (in `)	Number of Persons
Below 500	4
501 - 1000	10
1001 - 1500	14
1501 - 2000	18
2001 - 2500	24
2501 - 3000	20
Above - 3001	25

Q11. What is frequency distribution? Explain the formation of discrete and continuous frequency distribution.

Ans :

Frequency Distribution

Frequency distribution refers to the tabular arrangement of data, when arranged into groups or categories according to conveniently established divisions of the range of the observations. In a frequency distribution, raw data is expressed in distinct groups called 'classes'. The number of observations that fall into each of these classes is known as 'frequency'. Thus, a frequency distribution has two parts, the left part represents the classes and the right part represents the fequencies.

Arrangement of raw data in the form of row or column is called data array. When data is described by a continuous variable, it is called continuous data and when described using discrete variable it is called discrete data.

The following are the two examples of discrete and continuous frequency distributions.

No. of Students	No. of Colleges
100	15
120	25
140	32
160	50
180	27
200	25

Table (1): Discrete Frequency Distribution

Age (In years)	No. of Students
10-15	12
15-20	15
20-25	24
25-30	30
30-35	20
40-45	22

Table (2): Continuous Frequency Distribution

Construction of a Discrete Frequency Distribution

The process of preparing frequency distribution is very simple. For discrete data, place all the possible variables , values in ascending order in one column and then prepare another column for 'Tally marks' (|) to count the number of marks. A block of five marks is prepared and some space is left in between the blocks. The frequency column refers to the number of "Tally marks", a particular class will contain. The construction of a discrete frequency distribution is illustrated as follows.

Consider a sample study in which 40 families were surveyed to find the number of children in each family. The data obtained are,

3	2	2	1	3	4	2	1	3	4
5	0	2	1	2	3	3	2	1	1
2	3	0	3	2	1	4	3	5	5
4	3	6	5	4	3	1	0	6	5

'Tally marks' are used to condense this data into a discrete frequency distribution as shown below.

No. of Children	No. of Families	Frequency
0	III	3
1	𝍸 II	7
2	𝍸 III	8
3	𝍸 𝍸	10
4	𝍸	5
5	𝍸	5
6	II	2
Total		40

Table: Discrete Frequency Distribution

Construction of a Continuous Frequency Distribution

Before constructing the frequency distribution for continuous data, it is necessary to identify some of the important terms that are used frequently.

- **Class Limits**

  The class limits represent the smallest and the largest values to be included in a class. The lowest and highest boundaries of a class, are known as the lower and upper limits of the class respectively.

  For example, in a class among the values ranging between 70-79, 70 is the lower limit and 79 is the upper limit or we can say that there can be no value which is less than 70 and greater than 79.

- **Class Intervals**

  The class interval represents the width (span or size) of a class. The width may be determined by subtracting the lower limit of one class from the lower limit of the following class (alternatively successive upper limits may be used).

  For example, if the two classes are 20 – 30 and 30 – 40, the width of the class interval would be the difference between the two successive lower limits i.e., 30 – 20 = 10 or the difference between the upper limit and lower limit of the same class, i.e., 30 – 20 = 10.

- **Class Midpoint**

  Midpoint of a class is defined as the average of the lower and upper limits of a class. The value thus obtained lies between the lower and upper class limits. For example the midpoint of a class 10-20 is,

$$\frac{10+20}{2}=\frac{30}{2}=15$$

- **Class Frequency**

  The number of observations that fall into each of the class is known as the class frequency. For example, the class frequency for the class interval 20 – 30 is ‘10’.

Types of Class Interval

There are different ways in which limits of a class interval can be expressed such as,

1. Exclusive Method

The class intervals are arranged in such a way that the upper limit of one class is the lower limit of the next class. For example, a table below represents an exclusive method for class intervals.

Sales (` in thousands)	No. of Companies
30 - 40	22
40 - 50	29
50 - 60	35
60 - 70	42
70 - 80	14
80 - 90	18
90 - 100	27
100 - 110	61

Table: Exclusive Method for Class Interval

In the above example, there are 22 companies whose sales are between ` 30,000 and ` 39,999. A company with sales of exactly ` 40,000 would be included in the next class, i.e., 40-50.

Therefore in the exclusive method, it is always presumed that the upper limit is excluded.

2. Inclusive Method

In this method, the upper limit of one class is included in the class itself. For example, a table below represents the inclusive method of class intervals.

Sales (` in thousands)	No. of Companies
30 - 39	22
40 - 49	29
50 - 59	35
60 - 69	42
70 - 79	14
80 - 89	18
90 - 99	27
100 - 109	16

Table: Inclusive Method for Class Interval

Here, there are 22 companies whose sales are between ` 30,000 and ` 39,000. A company whose sales are exacth ` 40,000 would be included in the next class. Therefore in the inclusive method, it is presumed that the upper limit is included.

3.3 Diagrammatic and Graphical Representation of Data

Q12. Define diagram and diagramatical represenation. Explain general rules for constructing diagrams.

Ans : **(Imp.)**

(i) Diagrams

Diagram is a visual presentation of statistical information. The pictorial presentation helps in proper understanding the data . Diagrams are of different types like pie diagram, rectangles, lines, pictures and maps.

(ii) Diagrammatic Representation of Data

Apart from classification and tabulation methods of efficiently presenting the statistical data, the statistical data can also be presented in a convincing, appealing and easily understandable form using the diagrammatic representation (i.e., making use of diagrams).

General Rules For Constructing Diagrams

The following general rules should be observed while constructing diagrams:

1. Title

Every diagram must be given a suitable title. The title should convey in as few words as possible the main idea that the diagrams intend i portray. However, the brevity should not be secured at the cost of clarity or omission of essential details. The title may be given either at the top of the diagram or below it.

2. Proportion between Width and Height

A proper proportion between the height and width of the diagram should be maintained. If either the height and width is too short or too long in proportion, the diagram would give an ugly look. While there are no fixed rules about the dimensions, a convenient standard as suggested by Lutz in the book entitled "Graphic Presentation" may be adopted for general use. It is known as "Root-two" that is, a ratio of 1 (short side) to 1.414 (long side). Modifications may, no doubt, be made to accommodate a diagram in the space available.

3. Selection of Scale

The scale showing the values should be in even numbers or in multiples of five or ten, *e.g.,* 25, 50, 75, or 20, 40, 60. Odd values like 1, 3, 5, 7 should be avoided. Again no rigid rules can be laid down about the number of rulings on the amount scale, but ordinarily it should not exceed five. The scale should also specify the size of the unit and what it represents; for example, "million tonnes", "number of persons in thousands", "units produced in lakhs", etc. All lettering should be easily readable without turning the chart up and down.

4. Footnotes

In order to clarify certain points about the diagram, footnote may be given at the bottom of the diagram.

5. Index

An index illustrating different types of lines or different shades, colours should be given so that the reader can easily make out the meaning of the diagram.

6. Neatness and Cleanliness

Diagrams should be absolutely neat and clean.

7. Simplicity

Diagrams should be as simple as possible so that the reader can understand their meaning clearly and easily. For the sake of simplicity, it is important that too much material should not be loaded in a single diagram otherwise it may become too confusing and prove worthless. Several simple charts are often better and more effective than one or two complex ones which may present the same material in a confusing way.

Q13. Explain advantages and disadvantages of diagrams.

Ans :

Advantages

1. Diagrams give a clear picture of the data. Data presented with the help of diagrams can be understood and grasp even by a common man in a very short time.
2. Diagrammatical representation of data can be used universally at any place.
3. Diagrammatical representation not only saves time and energy but also is economical.
4. By using diagrammatic representation, data can be condensed.
5. By diagrammatic representation, comparison between data can be made without actually computing the statistical measures.
6. Diagrams are more impressive, attractive and fascinating compared to any other form of representation.
7. More information can be obtained using diagrammatic representation when compared to tabular representation of data.
8. Diagrams have an appealing effect to the user and are more easily remembered than any other representation.

Disadvantages

Some of the disadvantages of diagrammatic presentation are,

1. Diagrams do not show exact values.
2. Drawing diagrams is a difficult task.
3. Skills are required to present the data in a diagrammatic form.
4. Diagrammatic presentation of data reveals limited facts.
5. Diagrams are not substitute for tabular presentation.
6. Diagrams give only rough idea. Analyzing the data in detail is not possible through diagrams.
7. Diagrams drawn through false data give false idea.

Q14. What do you mean by graphical represen tation of data ? State the rules for graphing. What are the advantages and disadvantages of graphs ?

Ans : **(Aug.-21)**

Graphical Representation of Data

The graphical representation refers to the way of representing the data with the help of graphs. Graphs play a significant role in representation of data related to time series. The frequency distributions can also be efficiently represented using graphs. The categorical and geographical data can be best represented using a diagrammatic representation. Similarly, graphical representation proves to be the best approach for representing the data related to time series and frequency distribution.

Rules for Graphing

The various rules for constructing graphs are,

1. Title

A heading depicting the matter and contents of the data must be provided as title for all the graphical representations. The title must be clear, simple and precise.

2. Organization of the Graph

Each graph must be structurally drawn in an attractive manner. The portion of axis must be chosen and drawn neatly. We must represent independent and dependent parameters on X and Y axis respectively.

3. Scale

The scale selected must satisfy all the values to be plotted on the graph.

4. Index

The index must be provided to show the scale of X and Y axis. The various lines drawn in the graph must be defined clearly in the index.

5. Source of Data

The source of data gives the information about the data retrieval and is mentioned at the bottom of the graph.

6. False Base Line

The vertical scale must start with zero for representing the given variable effectively. When the fluctuations are maximum and the starting values are very far from zero, a false base line is used to break the vertical scale and to plot the values.

To draw a false base line, the vertical scale is broken into two parts and the values of the dependent variable from zero to the lowest value are omitted by drawing two zig-zag horizontal lines above the base line i.e., X-axis.

7. Line Designs

In case of representing several variables on the same graph for comparison, we may differentiate between them by using dotted lines, dash-dotted lines, broken lines, thick lines and thin lines. An index must also be provided to clarify the meanings of the lines.

Advantages of Graphs

Some of the advantages of graphical presentation are,

(i) Graphs represent the data in a simple manner which is easy to understand by the users.

(ii) The curves drawn on graphs are helpful in determining the values which are not available.

(iii) Easy comparisons can be made, if the data is presented in a graphical manner.

(iv) Graphs are highly attractive.

(v) Graphs can be used for further data analysis.

(vi) Graphs are useful in extrapolation and intrapolation methods.

Disadvantages of Graphs

There are certain disadvantages of Graphs. They are,

(i) One should possess skills to understand the graphs.

(ii) Graphs are not precise when compared to tables.

(iii) Constructing a graph is a time consuming process.

(iv) At times, graphs create problems.

(v) Qraphs provide restricted information.

Q15. List and explain the different ways of representing frequency distribution using graphs.

Ans : **(Nov.-20.)**

The graphical method of representing the frequency distribution is very useful. The frequency graphs provide a clear view, which can be easily understood and remembered by the users as compared to the tabulated data. The frequency distribution can be presented graphically in any of the following ways,

1. Histogram
2. Frequency polygon
3. Smoothed frequency curve
4. Cumulative frequency curves or Ogives.

1. Histogram

One of the most commonly used and easily understood methods for graphical representation of frequency distribution data is called 'histogram'. It is also known as 'column diagram', depicting the class frequencies in a frequency distribution by vertical adjacent rectangles.

2. Frequency Polygon

A frequency polygon is a graph representing frequency distribution. It has more than four sides and is very helpful in comparing several frequency distributions plotted on the same graph. This feature of frequency polygon is an advantage over histogram.

The frequency polygon can be constructed in two ways. They are,

(a) Using histograms

(b) Using midpoints.

(a) Using Histograms

A histogram is drawn by using the given data. Identify the midpoints of the upper horizontal side of the rectangles. Join the midpoints by a line. The figure thus obtained is known as ' frequency polygon'. The area of the frequency polygon is approximated by joining the ends of the polygon with the base line and by assuming that both the frequencies of the class before the first class and the class after the last class are zero.

(b) Using Midpoints

Calculate the midpoints of the various class intervals. Plot the midpoints and their corresponding frequencies on the graph. Join these points by a straight line. The figure thus formed is called 'frequency polygon'.

3. Smoothed Frequency Curve

A smoothed frequency curve is a free hand curve obtained from the various points of the polygon. The area included under the curve is approximately the same as that of polygon.

It is another way of presenting data where only frequency distributions based on samples should be smoothed. The continuous series thus generated is smoothed. The smoothed frequency curve is constructed by first drawing a histogram, plotting the frequencies at the midpoints of class intervals that results in a polygon. Finally, it is smoothed to produce a smoothed frequency curve. Therefore, the total area under the curve is equal to the area under the original histogram or polygon.

4. Cumulative Frequency Curves or Ogives

The cumulative frequency curve also known as Ogive is a graph of cumulative frequency distribution. These cumulative frequencies are obtained by adding the frequencies and listing them in a tabular form.

The two techniques for constructing an Ogive are,

(a) More than technique

(b) Less than technique.

(a) More than Technique

Here, more than cumulative frequencies are plotted on Y-axis and the lower limit of the class interval on X-axis. The points are then joined to obtain a smooth free hand curve called the 'More than Ogive'.

(b) Less than Technique

In this technique, less than cumulative frequencies are plotted on Y-axis and the upper limit of the class interval on X-axis. The points are-then joined so as to form a smooth free hand curve called the 'Less than Ogive.

Q16. Differentiate between diagrams and graphs.

Ans :

The differences between Diagrams and Graphs are as follows,

Diagrams	Graphs
1. Diagrams are usually constructed on a plain sheet and they are useful for comparison purpose.	1. Graphs are constructed on a special sheet known as graph paper and they are used for studying the relationship between the variables
2. In Diagrammatic presentation, data is presented by using bars, pie charts, circles, cubes etc.	2. In Graphical presentation, data is presented through lines or points of various types like dots, dashes, dot-dash and so on.
3. Diagrams give aproximate idea and are less accurate.	3. Graphs provide reliable, accurate and authentic information.
4. Compared to graphs, constructing diagrams is a difficult task.	4. They are easy to construct compared to diagrams.
5. Geographical and categorical data can be presented through diagrams.	5. Data concerning time series and frequency distribution can be presented through graphs.

3.3.1 One Dimensional, Two Dimensional and Three-Dimensional Diagrams and Graphs

Q17. What do you mean by one-dimensional diagram? What are the commonly used one-dimensional diagrams?

Ans : **(Imp.)**

One Dimensional Diagram

The simple bar chart with large number of variations and which is particularly appropriate for comparing the magnitude (or size) of co-ordinate items or of parts of a total. The basis of comparison in the bar is "linear" or one-dimensional in nature.

- Calvir F. Schmid

One dimensional diagrams are usually considered because it constitutes only the length of the bar (i.e., height) and excludes the width of the bar. The diagram is not effected by the width or thickness of the bar but utmost care must be taken to see that the diagrams thickness is not very high that it would appear as a two-dimensional diagram.

Thus, these diagrams are not only useful in diagrammatical presentation of frequency distributions but also in understanding shape of the distribution and their features.

Commonly used One Dimensional Diagrams

The two commonly used one-dimensional diagrams are,

1. Line Diagram

The line diagram is the simplest kind of diagram where the frequencies of the discrete variable are shown. The diagram is constructed by taking the variable on the X-axis and observed frequencies on Taxis. The resultant diagram comprises of vertical lines whose lengths are proportional to the frequencies.

Example

Mutant strains of Ergot (CJaviceps Purpura) obtained by induced mutagenesis technique are shown in the table

Mutant strains of Ergot	Frequency	Relative
1. C. Purpurea A Tcc 15383	50	0.38
2. C. Purpurea A Tcc 20103	44	0.33
3. C. Purpurea A Tcc 20019	36	0.27
Total	**130**	**0.98**

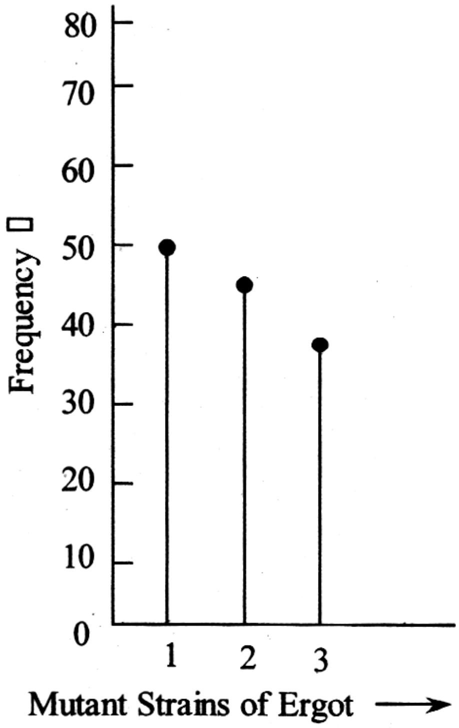


Fig: Line Diagram

2. Bar Diagram

These diagrams are very simple and easy to construct. They consider only the length of the bars but not their widths. These type of diagrams represents the data with the help of thick lines or bars.

Points to Remember and Rules to be Followed While Constructing Bar Diagrams

i) Maintain uniformity in the width of the bars.

(ii) Represent the bars either vertically or horizontally. Easy comparison and better appearance is achieved with the help of vertical arrangement of bars,

(iii) Maintain uniformity in the gaps between the bars.

(iv) The bar diagrams must be drawn in such a way that every bar has its respective figure at the top. By doing so, we can avoid confusion among readers and make easy comparison of various objects.

Merits of Bar Diagram

Some of the merits of bar diagrams are as follows,

(i) Every individual irrespective of whether they are accustomed to reading charts or not, can understand bar diagrams.

(ii) Mass data can be compared effectively by using bar diagrams.

(iii) Compared to other diagrams, constructing bar diagrams is easy and simple.

Q18. Explain different types of bar diagram.

Ans : **(Imp.)**

Bar diagrams are of the following types :

(a) Simple bar diagrams

(b) Sub - divided bar diagrams

(c) Multiple bar diagrams

(d) Percentage bar diagrams

(e) Deviation bars.

(a) Simple Bar Diagrams

A simple bar diagram is used to represent only one variable. For example, the figures of sales, production, population, etc., for various years may be shown by means of a simple bar diagram. Since these are of the same width and only the length varies, it becomes very easy for the reader to study the relationship. Simple bar diagrams are very popular in practice. This can be either vertical or horizontal. In practice, vertical bars are more popular. However, an important limitation of such diagrams is that they can present only one classification or one category of data. For example, while presenting the population for the last five decades, one can only depict the total population in the simple bar diagrams, and not its sex-wise distribution.

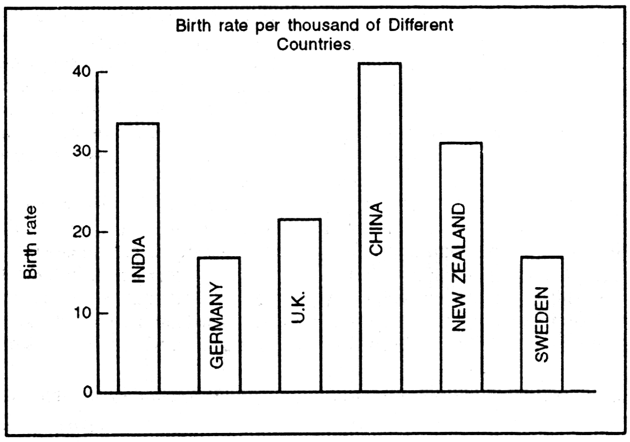
Example

Following table gives the birth rate per thousand of different countries over a certain period :

Country	Birth Rate	Country	Birth Rate
India	33	China	40
Germany	16	New Zaland	30
U.K.	20	Sweden	15

Represent the above data by a suitable diagram.

Sol :



(b) Sub-divided Bar Diagrams

In a sub-divided bar diagram each bar representing the magnitude of a given phenomenon is further sub-divided in its various components. Each component occupies a part of the bar proportional to its share in the total.

For example, the number of students in various courses for B.Com., M.Com., BA. and M.A. in various colleges may be represented by a sub-divided bar diagram. While constructing such a diagram, the various components in each bar should be kept in the same order. A common and helpful arrangement is that of presenting each bar in the order of magnitude from the largest component at the base of the bar to the smallest at the end. To distinguish between the different components, it is useful to use different shades or colours, index or key should be given explaining these differences.

Sub-divided bar diagrams should not be used where the number of components is more than 10 or 12, for, in that case, the diagram will be cverloaded with information which cannot be easily compared and understood.

The component bar diagram can be used to represent either the absolute data or distribution ratios such as percentage distributions. It is, in fact, an excellent method for presenting a set of distribution ratios diagrammatically. The sub-divided bar diagrams can be constructed both on horizontal and vertical bases.

Example : The growth of production of Rsh for the period 1991-92 to 1997-98 is given below:

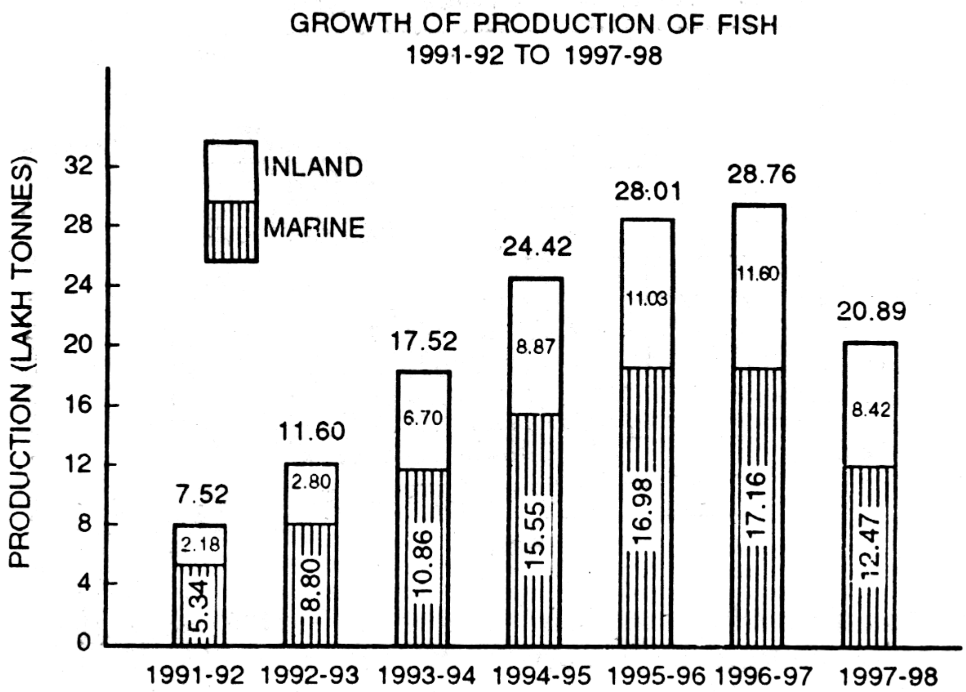
(lakh tonnes)

Year	Marine	Inland	Total
1991 - 92	5.34	2.18	7.52
1992 - 93	8.80	2.80	11.60
1993 - 94	10.86	6.70	17.56
1994 - 95	15.55	8.87	24.42
1995 - 96	16.98	11.03	28.01
1996 - 97	17.16	11.60	28.76
1997 - 98	12.47	8.42	20.89

Represent the data by a suitable diagram.

Sol :

The above data can be represented by a suh-divided bar diagram (drawn on a vertical base. to



(c) Multiple Bars

In a multiple bar diagram two or more sets of interrelated data are represented. The technique of drawing such a diagram is the same as that of simple bar diagram. The only difference is that since more than one phenomenon is represented, different shades, colours, dots or crosses are used to distinguish between the bars. Wherever a comparison between two or more related variables is to be made, multiple bar diagram should be preferred.

Example

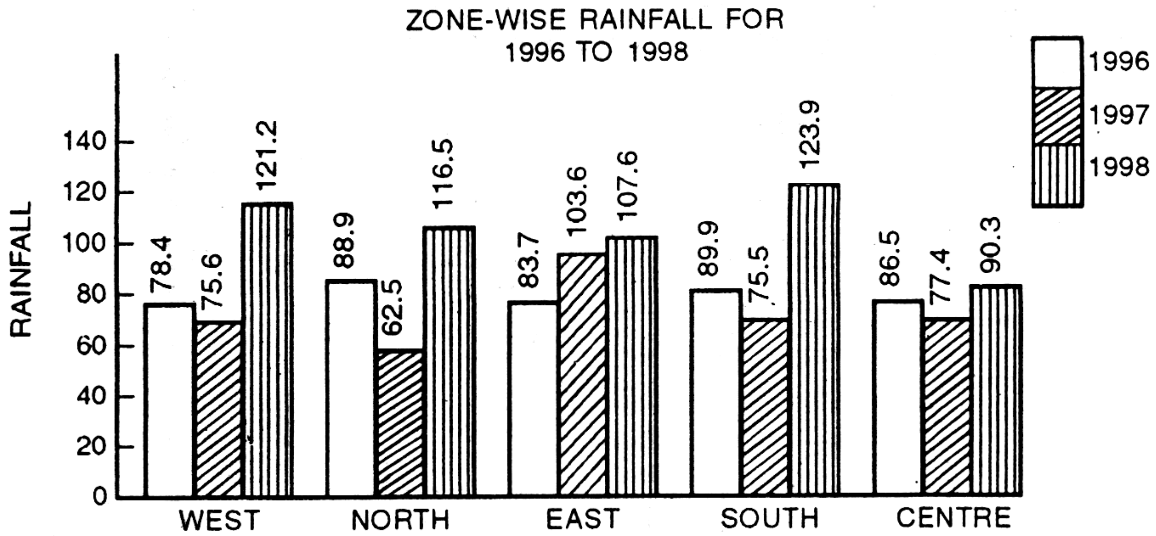
The regional rainfall indices during the year 1997 to 1998 are given below :

Zone

Year	West	North	East	South	Centre
1996	78.4	88.9	83.7	89.9	86.5
1997	75.6	62.5	103.6	75.5	77.4
1998	121.2	116.5	107.6	123.9	90.3

Represent the data by a multiple bar diagram.

Sol :



(d) Percentage Bars

Percentage bars are particularly useful in statistical work Which requires the portrayal of relative changes in data. When such diagrams are prepared, the length of the bars is kept equal to 100 and segments are cut in these bars to represent the components (percentages) of an aggregate.

Example

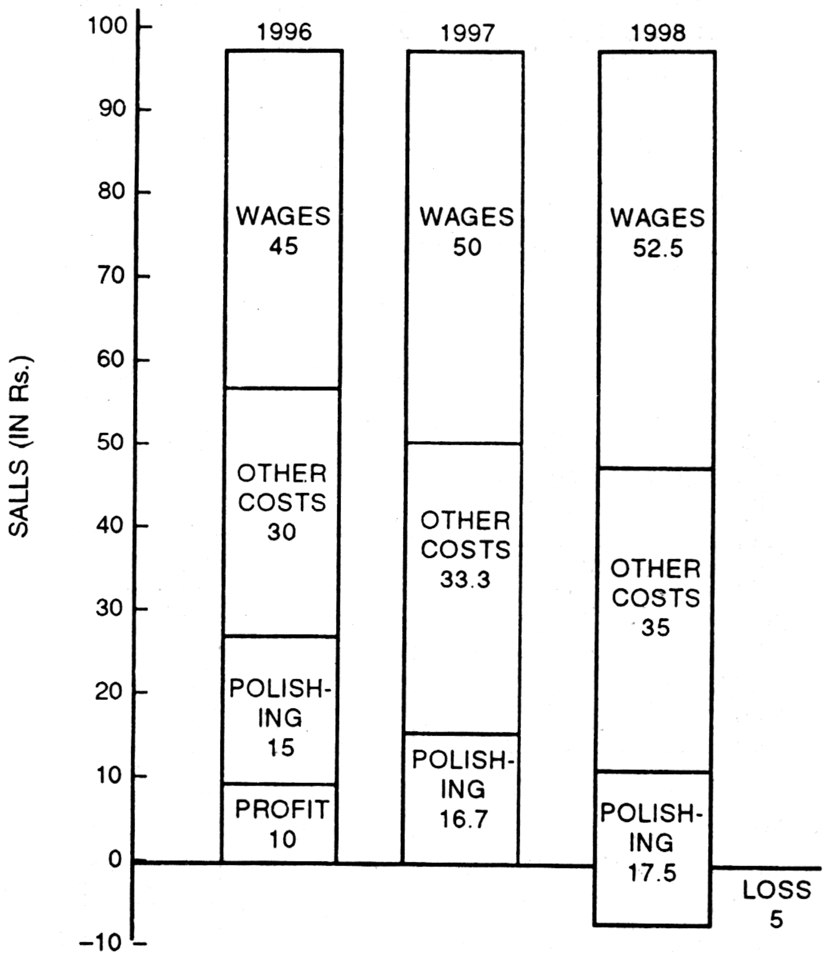
Represent the following by sub-divided bar diagram drawn on the percentage basis :

	Particulars	1996	1997	1998
1.	Cost per chair			
	(a) Wages	9	15	21
	(b) Other costs	6	10	14
	(c) Polishing	3	5	7
	Total	18	30	42
2.	Proceeds per chair	20	30	40
3.	Profit (+)	+2	–	– 2
	Loss (-)			

Sol :

Take the sale price per chair as Rs. 100 and express the other figures in , oercentages. The percentages so obtained are given below:

Particulars	1996	1997	1998
Wages	45.0	50.0	52.5
Other costs	30.0	33.3	35.0
Polishing	15.0	16.7	17.5
Total costs	90.0	100.0	105.0
Sale price	100.0	100.0	100.0
Profit or loss	+10.0	—	-5.0



(e) Deviation Bars

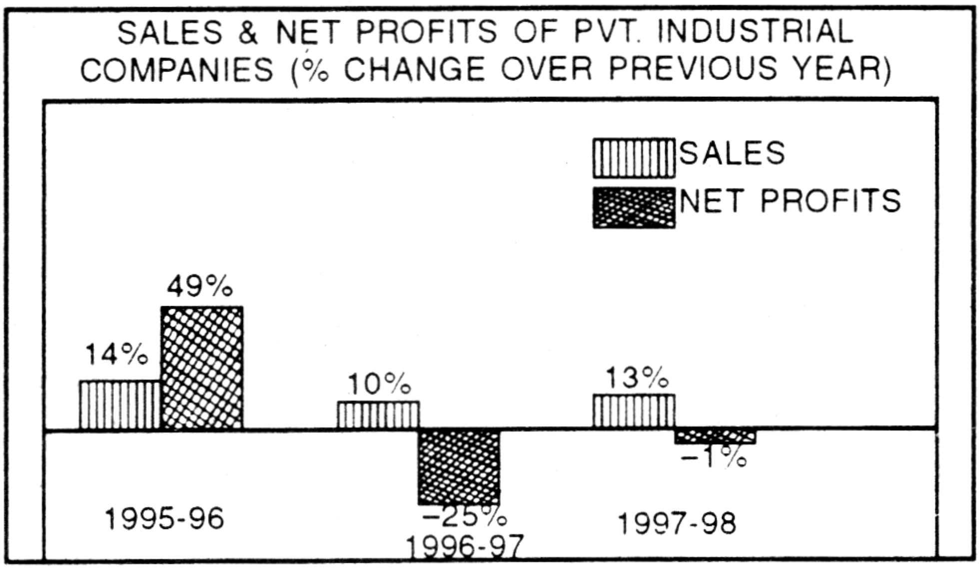
Deviation bars are popularly used for representing net quantities - excess or deficit, ie., net profit, net loss, net exports or unports, etc. Such bars can have both positive and negative values. Positive values are shown above the base line and negative values below it. The following illustration would explain this type of diagram:

Example :

Present the following data by a suitable diagram showing the Sales and Net Profits of private industrial companies:

Year	Sales	Net Profits
1995 - 96	14%	49%
1996 - 97	10%	– 25%
1997 - 98	13%	–1 %

Sol :



Q19. Explain briefly two dimensional and three dimensional diagrams.

Ans : **(Nov.-20)**

Two Dimensional Diagrams

Two dimensional (2D) diagrams are also referred as 'surface diagrams' or 'area diagrams'. These diagrams are constructed with respect to length and width of the given data,

Types of Two Dimensional Diagrams

Some of the important 2-D diagrams are listed below.

1. Squares
2. Rectangles
3. Circles.

1. Squares

These type of 2-D diagrams are very much useful for representing data where the values of the items vary widely in magnitude.

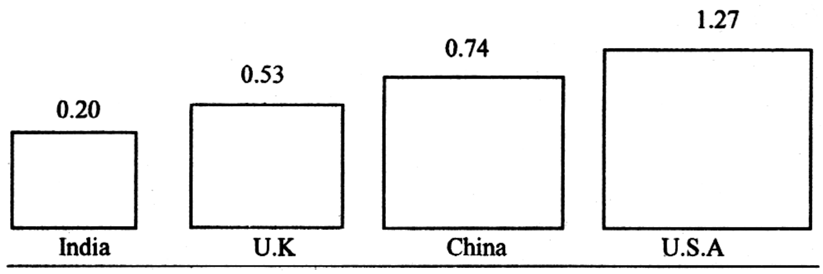


Fig : Square Diagram Showing Coal Production Rectangles

2. Rectangles

Rectangles are quite popular for representing the data. These involve comparison between two or more items, where every item is subdivided into several components.

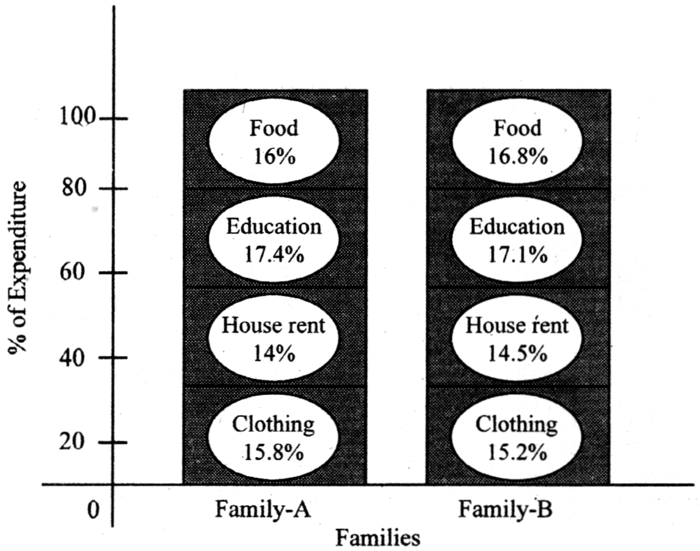


Fig : Rectangle Diagram Showing % of Expenditure fay Two Families

3. Circles

Circles can be used when there is too large data for comparison and in such cases it is very difficult to construct squares. Circles are difficult to compare and are not very popular. Hence when it is possible to represent data by bars, we should avoid using circles.

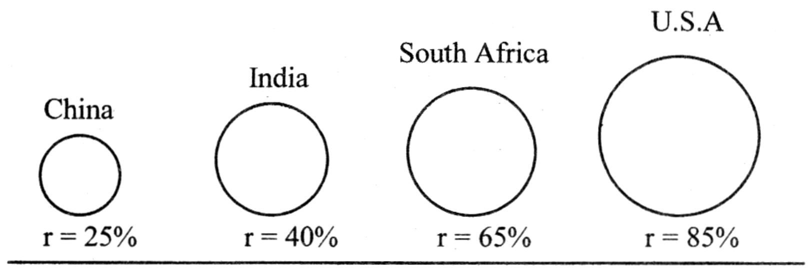


Fig : Circle Diagram Showing Land Area

Pie Diagram

It is also known as circle diagram and ranks very high with respect to its popularity. It partitions the whole commodity into its various component parts. This diagram represents the distribution of things related to a particular element and facilitate comparison between various components of it.

Steps for Constructing Pie Diagram

The steps involved in constructing pie diagram are,

1. Plot a circle of appropriate radius with a total of 360°.
2. Convert different commodities into degrees by multiplying the given percentages with 360 i.e.. 360 x Percentage value of the commodity.
3. Arrange sectors according to their size with the largest at the top and others in sequence running clockwise.
4. Points on a circle representing the size of each sector are finally measured with the help of a protractor.

Example

Draw a suitable diagram to represent the following data relating to the production cost of manufacture.

Cost of material = ` 38,400

Cost of labour = ` 30,720

Direct expenses = ` 11,520

Overhead expenses = ` 15,360

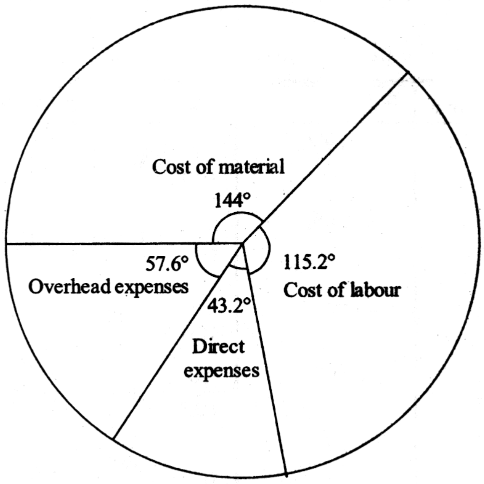
Sol :

Total Production Cost = Cost of Material + Cost of Labour + Direct Expenses + Overhead Expenses

= 38,400 + 30,720 + 11,520 + 15,360 = 96,000.

Computations for construction of pie diagram for the given data are as follows,

Item of Expenditure	Expenditure in (`)	Expenditure in Percentage (%)	Expenditure in Degree (°)
Cost of material	38,400	$\frac{38400 \times 100}{96000} = 40\%$	$\frac{40 \times 360}{100} = 144^{o}$
Cost of labour	30,720	$\frac{30720 \times 100}{96000} = 32\%$	$\frac{32 \times 360}{100} = 115.2^{o}$
Direct expenses	11,520	$\frac{11520 \times 100}{96000} = 12\%$	$\frac{12 \times 360}{100} = 43.2^{o}$
Overhead expenses	15,360	$\frac{15360 \times 100}{96000} = 16\%$	$\frac{16 \times 360}{100} = 57.6^{o}$
Total	**96,000**	**100%**	**360°**



Limitations of Pie Diagram

Following are the limitations of pie diagram,

1. Interpreting and reading pie diagram is ineffective when compared to bar diagrams.
2. When series of data is divided into large number of components then it becomes difficult to represent such data through pie diagram.

Three Dimensional (3-D) Diagrams

Three-dimensional diagrams are also referred as 'volume diagrams'. These diagrams are constructed with respect to length, width and height of the given data.

Three-dimensional diagrams are useful in situations where the range of difference between the minimum value and the maximum value is very large. Some of the examples of three dimensional diagrams are cubes, cylinders and spheres.

Limitations of Three Dimensional Diagrams

Following are the limitations of three dimensional diagrams,

1. These diagrams are very difficult to interpret and understand.
2. 3-D diagrams are not very much useful for statistical presentation.

Q20. Explain the presentation of data by using graphs.

Ans :

Frequency distributions can be described with the help of graphs. Graph is more distinct in comparison to diagram and it is comparatively easier to draw graph than diagram. Data of frequency distributions and time series data are generally represented by graphs. Generally diagrams are drawn to make comparative study. But diagrams do not depict the relationships among the variables to be studied. With the help of graphs the relationships among the variables can be studied. The importance of graph is more than that of diagram in statistical analysis.

In the practical field various types of graphs are used. These graphs can broadly be divided into two groups:

a) Graphs of frequency distribution

b) Graphs of time series

Graphs of frequency distribution: In order to make frequency distribution easily understandable three types of graphs are usually drawn. These are:

(i) Histogram or column diagram

(ii) Frequency Polygon

(i) Histogram or column diagram

The graph by which the frequencies of various class intervals of a frequency distribution with the help of adjacent vertical rectangles are shown is called a histogram or column diagram. First of all, the actual class-intervals are to be marked on the x-axis choosing a suitable scale. Then taking these as bases, rectangles are to be drawn continuously on this basis. When the class intervals are equal then the heights of the rectangles will be proportional to the frequencies of the corresponding class intervals, but if the class intervals are not equal then the heights of the rectangles will be proportional to the ratios between the corresponding class frequencies and the lengths of the class intervals.

Example :

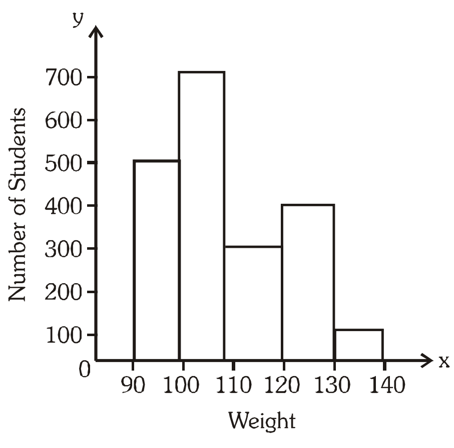
Draw a histogram to represent the following distribution:

(Frequency distribution of weights of 2000 students)

Weights (in lbs)	90-100	100-110	110-120	120-130	130-140
No. of students	500	700	300	400	100

Sol :

The class intervals of the above frequency distribution are of equal length. First of all, the class intervals are marked on the X-axis and then rectangles of heights proportional to the frequencies of the class intervals are drawn on them.



ii) Frequency Polygon

The graph obtained by joining the mid-points of the upper horizontal sides of the adjacent rectangles of a histogram by line segments is called a frequency polygon. Frequency polygon can be drawn directly without obtaining indirectly from histogram. For this we are to take the mid-points of the class intervals on the X-axis and the frequencies of the class intervals are to be plotted against the corresponding class mid-points along the Y-axis. Then these plotted points are to be joined by line segments. The end points of the graph so obtained are to be joined with the mid-points of the two frequency less class intervals which preceed and succeed the first and the last class intervals respectively. The graph obtained thereby is the desired frequency polygon.

In case the class intervals of a frequency distribution are not all equal then frequency polygon is usually not drawn.

Example :

The following data show the number of accidents sustained by 313 drivers of a public utility company over a period of 5 years.

Number of accidents :	0	1	2	3	4	5	6	7	8	9	10	11
Number of drivers	80	44	68	41	25	20	13	7	5	4	3	2

Draw the frequency polygon.

Sol :.

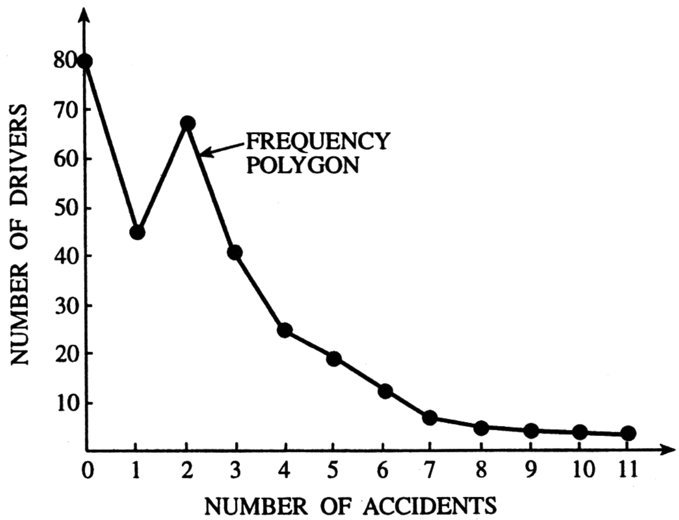


Fig.: Frequency Polygon of Number of Accidents

PROBLEMS

4. **Draw a pie diagram for the following data of Sixth Five - Year Plan Public Sector outlays :**

Agriculture and Rural Development	**12.9%**
Irrigation, etc.	**12.5%**
Energy	**27.2%**
Industry and Minerals	**15.4%**
Transport, Communication, etc.	**15.9%**
Social Services and Others	**16.1%**

Sol :

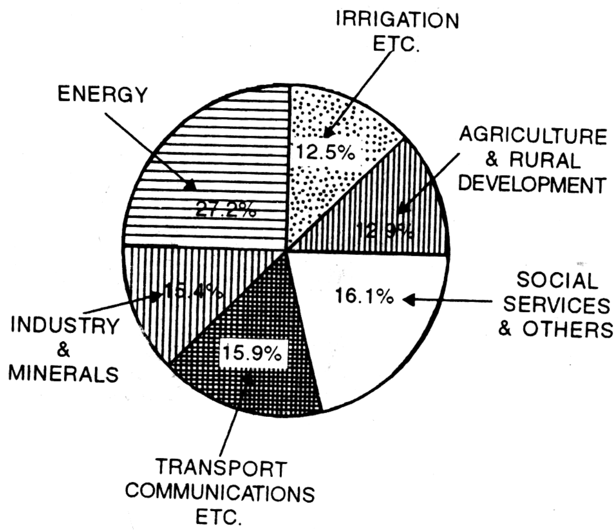
The angle at the centre is given by

$$\frac{\text{Percentage outlay}}{100} \times 360 = \text{Percentage outlay} \times 3.6$$

Computation for PIE Diagram

Sector	Percentage	Angle outlay
Agriculture and Rural Developement	12.9	12.9 × 3.6 = 46°
Irrigation, etc.	12.5	12.5 × 3.6 = 45°
Energy	27.2	27.2 × 3.6 = 98°
Industry and Minerals	15.4	15.4 × 3.6 = 56°
Transport, Communication, etc.	15.9	15.9 × 3.6 = 57°
Social Services and others	16.1	16.1 × 3.6 = 58°
Total	**100.0**	**360°**

Now a circle shall be drawn suited to the size of the paper and divided into 6 parts according to degrees of angle at the centre. (The angles have been arranged in descending order).



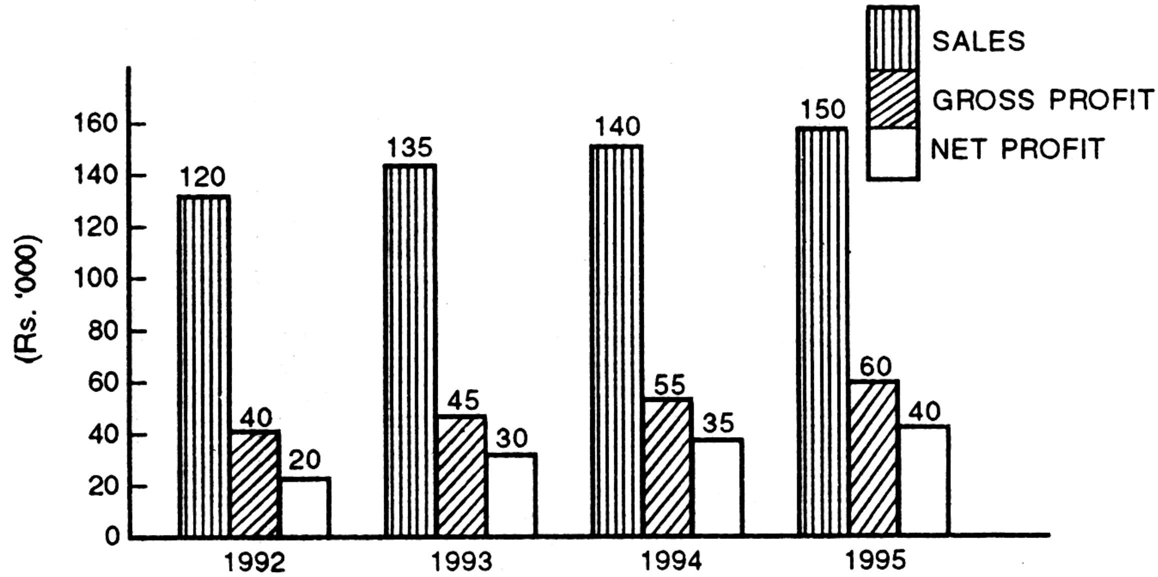
5. Draw a multiple bar diagram from the following data :

Year	Sales ('000 Rs.)	Gross Profit ('000 Rs.)	Net Profit ('000 Rs.)
1992	120	40	20
1993	135	45	30
1994	140	55	35
1995	150	60	40

Sol :

Sales, Gross Profits & Net Profits

(For the year 1992-95)



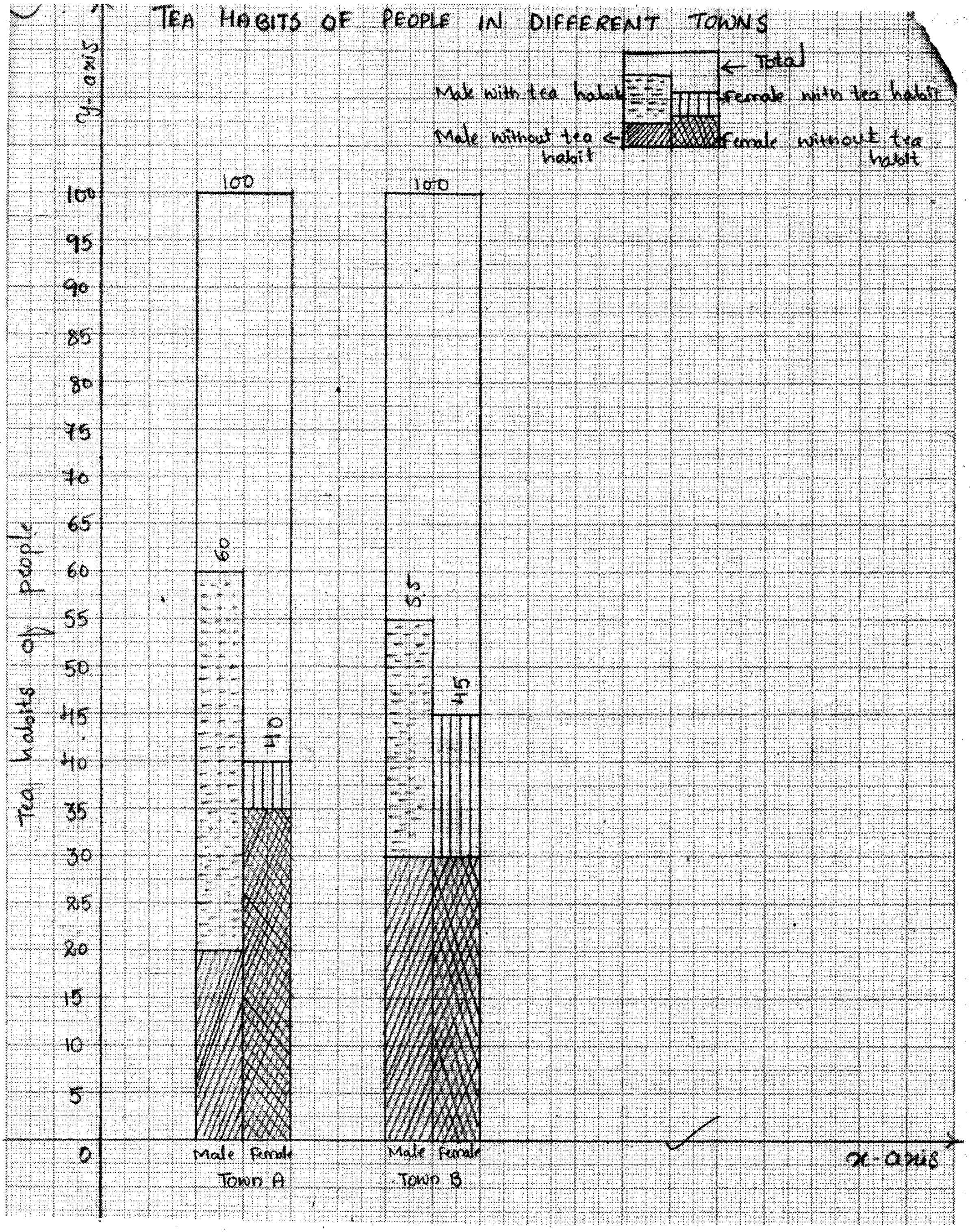
6. Represent the data shown in the following table by using suitable graphical method.

Department	Town A			Town B		
	Male	Female	Total	Male	Female	Total
No. of people with tea habit	40	5	45	25	15	40
No. of people with - out tea habit	20	35	55	30	30	60
Total	60	40	100	55	45	100

Sol :

Representation of Data by Sub - divided Multiple Bars Method

Department	Town A			Town B		
	Male	Female	Total	Male	Female	Total
No. of people with tea habit	40	5	45	25	15	40
No. of people with - out tea habit	20	35	55	30	30	60
Total	**60**	**40**	**100**	**55**	**45**	**100**

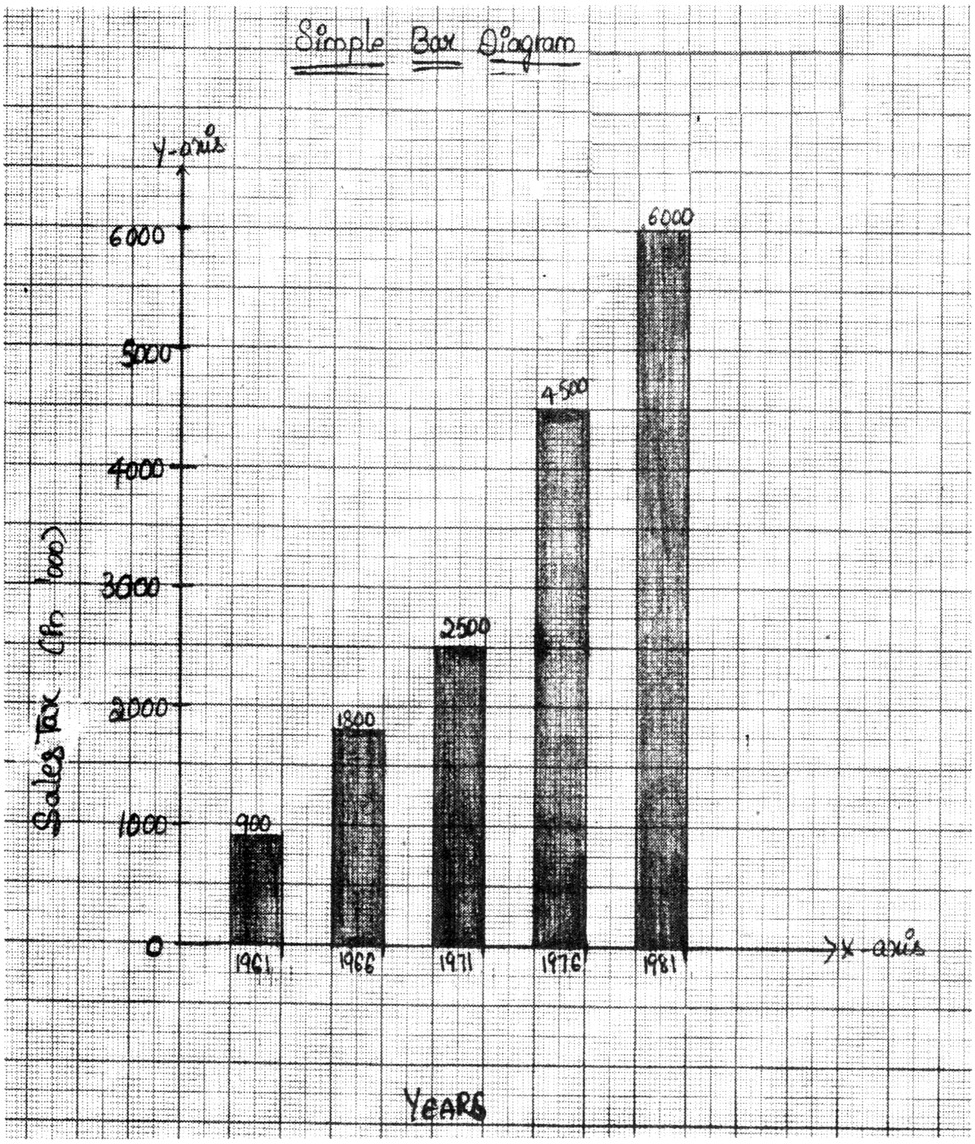

TEA HABITS OF PEOPLE IN DIFFERENT TOWNS
Total
Male with tea habit
Female with tea habit
Male without tea habit
Female without tea habit
y-axis
Tea habits of people
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
100
60
40
100
55
45
Male
Female
Town A
Male
Female
Town B
x-axis

Problem on Simple Bar Diagram :

7. Represent the below given data by a simple bar diagram :

Year	Sales Tax (in '000)
1961	900
1966	1800
1971	2500
1976	4500
1981	6000

Sol :



8. Draw pie diagram to represent the following data of proposed expenditure by a State Government for the year 1999 - 2000.

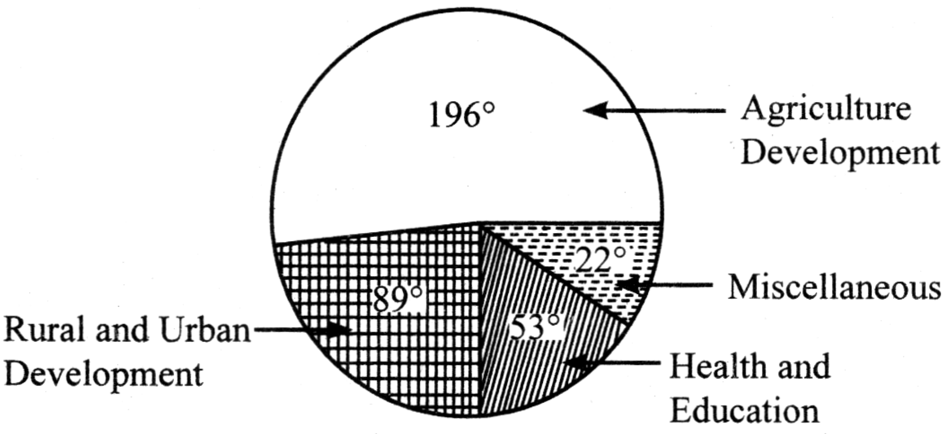
Items	Agriculture Development	Rural and Urban Development	Health and Education	Miscellaneous
Proposed Expenditure (in Million Rupees)	4400	2000	1200	500

Sol :

Calculations for Pie Chart

Items (a)	Proposed Expenditure (in Million Rupees) (b)	Angle at the Centre (c) = $\frac{(b)}{8100} \times 360^o$
Agriculture Development	4400	$\frac{4400}{8100} \times 360^o = 196^o$
Rural and Urban Development	2000	$\frac{2000}{8100} \times 360^o = 89^o$
Health and Education	1200	$\frac{1200}{8100} \times 360^o = 53^o$
Miscellaneous	500	$\frac{500}{8100} \times 360^o = 22^o$
	8100	**360°**

Pie diagram representing proposed expenditure by state government on different items for 1999 - 2000.



3.4 Small Sample Tests - t-Distribution - Properties and Applications

Q21. Define Small sample tests and t-distribution.

Ans : **(Nov.-21, Jan.-20)**

Small Sample test is one which consist of sample size less than (n<30) in small samples we can't assume the normal Distribution Approxmently it is denoted by "t" it is also called as "t" distibution in small sample test population size is not known in "t" test we assumed that the population from sample as been taken is normal.

Acceptance & Rejectence Criteria :

i) If $t_{cal} < t_{tab}$ Accept H_0

ii) If $t_{cal} > t_{tab}$ reject H_0

Here t_{cal} = calculated value

t_{tab} = tabulated value

If we take a very large number of small samples from a population and calculate the mean for each sample and then plot the frequency distribution of these means the resulting sampling distribution would be the **Student's t-distribution**.

The greatest contribution to the theory of small samples was made by **Sir William Gossett and R.A. Fisher**. Gossett published his discovery in 1905 under the pen name 'students' and its popularly known as t-test or students' t-distribution or students' distribution.

When the sample size is 30 or less and the population standard deviation is unknown, we can use the t-distribution.

The formula is, $t = \frac{(\bar{X} - \mu)}{S} \times \sqrt{n}$

Where S, or sample using n–1 as denominator then

$$t = \frac{(\bar{X} - \mu)}{S / \sqrt{n-1}}.$$

It should be noted that the only difference in the calculation of S in large and small samples is that whereas in the case of the former the sum of the squares of deviations of various items from the mean of $\Sigma(X - \bar{X})^2$ is divided by n–1. (the number of items) in case of small samples it is divided by n–1, which are the degrees of freedom.

The degrees of freedom in such problems is n–1 because one has the freedom to change only n–1 items as the last item has to be the difference between SX and sum of n–1 items. Thus, if there are 5 items in a sample with a total of 30. We can change only 4 items as we like. The last item will have to be 30 minus the sum of the remaining 4 items, whose values have been changed.

Q22. Explain the properties and applications of t-distribution.

Ans : **(Nov.-21)**

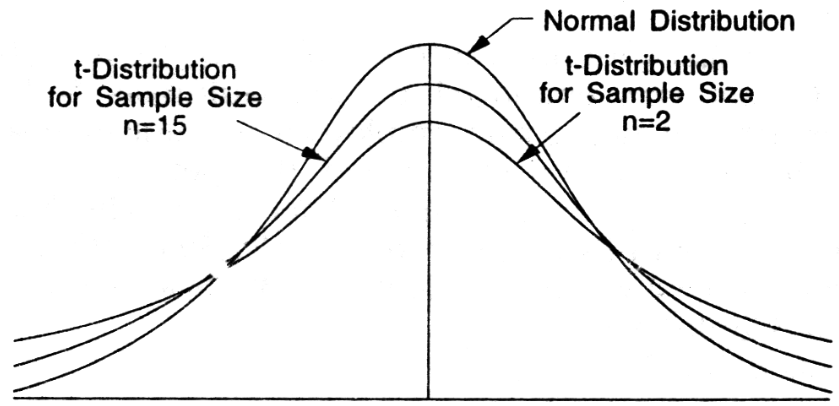
1. The variable t-distribution ranges from minus infinity to plus infinity.

2. The constant *c* is actually a function of *v* (pronounced as nu), so that for a particular value of *v*, the distribution of f(t) is completely specified. Thus f(t) is a family of functions, one for each value of *v*.

3. Like the standard normal distribution, the t-distribution is symmetrical and has a mean zero.

4. The variance of the t-distribution is greater than one, but approaches one as the number of degrees of freedom and, therefore, the sample size becomes large. Thus the variance of the t-distribution approaches the variance of the standard normal distribution as the sample size increases. It can be demonstrated that from an infinite number of degrees of freedom ($v = \infty$), the t-distribution and normal distribution are exactly equal. Hence there is a widely practised rule of thumb that samples of size n > 30 may be considered large and the standard normal distribution may appropriately be used as an approximation to t-distribution, where the latter is the theoretically correct functional form.

The following diagram compares one normal distribution with two t-distributions of different sample sizes :



Normal Distribution, *t*-Distribution for Sample Size *n*=15, and *t*-Distribution for Sample Size *n*=2.

The above diagram shows two important characteristics of t-distribution. First, a t-distribution is lower at the mean and higher at the tails than a normal distribution. Second, the t-distribution has proportionately greater area in its tails than the normal distribution. Interval widths from t-distributions are, therefore, wider than those based on the normal distribution.

The t- Table. The t-table given at the end is the probability integral of t-distribution. It gives, over a range of *v*. the probabilities of exceeding by chance value of t at different levels of significance. The t-distribution has a different value for each degree of freedom and when degrees of freedom are infinitely large, the t-distribution is equivalent to normal distribution and the probabilities shown in the normal distribution tables are applicable.

Applications of t-distribution

The following are some important applications of t-distribution,

(i) Test of hypothesis about the population mean.

(ii) Test of hypothesis about the difference between two means.

(iii) Test of hypothesis about the difference between two means with dependent samples.

(iv) Test of hypothesis about coefficient of correlation.

3.4.1 Testing for One and Two Means

Q23. Explain the procedure for testing of single mean.

Ans : **(Jan.-20)**

In determining whether the mean of a sample drawn from a normal population deviates significantly from a stated value (the hypothetical value of the populations mean), when variance of the population is unknown we calculate the statistic :

$$t = \frac{(\bar{X} - \mu)\sqrt{n}}{S}$$

where $\bar{X}$ = the mean of the sample

μ = the actual or hypothetical mean of the population

n = the sample size

S = the standard deviation of the sample

$$S = \sqrt{\frac{\Sigma\left(X-\bar{X}\right)^2}{n-1}}$$

Procedure of testing hypothesis :

Step 1 : Define null hypothesis (H_0)

Step 2 : Define Alternative hypothesis (H_1)

Step 3 : Deside level of significant

Step 4 : Test statistic

$$t = \frac{|\bar{x}-\mu|\sqrt{n}}{s}$$

$$s = \sqrt{\frac{\Sigma\left(x-\bar{x}\right)^2}{n-1}}$$

$\bar{x}$ = Mean of the sample

μ = Mean of the population

s = Standard deviation of sample

n = Sample size

PROBLEMS

9. A manufacturer of certain electric Bulbs claims that is bulbs have mean life 25 months with standard deviation 5 months a random sample of 6 such bulbs areas follows.

Life of Bulbs	1	2	3	4	5	6
Months	24	26	30	20	20	18

Can you regard the procedure clamied to be valid at 1% Los ($t_{0.01}$=4.032).

Sol :

Given that

Mean life (μ) = 25 months

Random sample (n) = 6 bulbs (6 < 30)

Standard deviation (σ) = 5 months

Level of Significant = 1%

Step 1 : Null hypothesis : (H_0)

There is no significant difference between mean life of bulbs

$$H_0 = \mu_1 = \mu_2$$

Step 2 : Alternative hypothesis : (H_1)

There is significant difference between mean life of bulbs

$$H_1 = \mu_1 \neq \mu_2$$

Step 3 : Level of significant

Los = 1%

$t_{0.01}$ = 4.032

Step 4 : Test Statistics

$$\text{"t"} \quad \frac{|\bar{X}-\mu|\sqrt{n}}{S}$$

$\bar{X}$ = Population mean

μ = Sample mean

n = Sample size

$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

X	$X-\bar{X}$	$(X-\bar{X})^2$
24	24 – 23 = 1	1
26	26 – 23 = 3	9
30	30 – 23 = 7	49
20	20 – 23 = – 3	9
20	20 – 23 = – 3	9
18	18 – 23 = – 5	25
		$\Sigma(x-\bar{x})^2 = 102$

$$\bar{x} = \frac{\Sigma x}{N}$$

$$= \frac{138}{6}$$

$\boxed{\overline{x} = 23}$

$$s = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n-1}}$$

$$s = \sqrt{\frac{102}{5}}$$

$$s = \sqrt{20.4}$$

$\boxed{s = 4.51}$

Substitute "S" value in "t"

$$\therefore\ t = \frac{|\overline{x}-\mu|}{s}\sqrt{n}$$

$$t = \frac{|23-25|\sqrt{6}}{4.51}$$

$$t_{cal} = \frac{|-2|\sqrt{6}}{4.51}$$

$$t_{cal} = \frac{(2)(2.44)}{4.51}$$

$$t_{cal} = \frac{4.88}{4.57}$$

$\boxed{t_{cal} = 1.08}$

$t_{cal} < t_{calse}$

1.08 < 4.032

$\therefore$ Accept the null hypothesis. (H_0)

$\therefore$ There is no significant different between mean life of bulbs.

10. A random sample of size 16 as 53 mean the sum of the squares of the deviation taken from the mean is 135. Can these sample regarded as taken from the population having 56 mean ? Obtain at 1% Los $t_{0.01} = 2.95$.

Sol :

Given that

Random sample size (n) = 16

Sample mean ($\overline{x}$) = 53

Mean sum of squares of the deviation = $\Sigma(x-\overline{x})^2$ = 135

Level of significance = 1% = $t_{0.01}$ = 135

Population mean (μ) = 56

Level of significance = 1% = $t_{0.01}$

Step (1) : The sample size is less than (30)

Hence it is considered as 't" test

Hypothesis :

1. **Null Hypothesis :** There is no significant difference b/w sample mean & population mean

$$H_0 = \overline{x} = \mu$$

Step 2 :

Alternative Hypothesis : There is significant difference b/w sample mean & population mean

$$H_1 = \overline{x} \neq \mu$$

Step 3 : Level of Significance

Loss = 1%

$t_{0.01}$ = 2.95

Degree of freedom (D.F) = n – 1

= 16 – 1

$\boxed{v = 15}$

Step 4 : Test statistic

$$t = \frac{|\overline{x}-\mu|\sqrt{n}}{s}$$

$$s = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n-1}}$$

$$s = \sqrt{\frac{135}{15}}$$

$\boxed{s = 3}$

Substitute "S" value in "t"

$$t = \frac{|53-56|\sqrt{16}}{3}$$

$$t = {}^{(3)}\!/_{3}\ (4)$$

$\boxed{t = 4}$

Step 5 : Acceptance & Rejectence Criteria

$t_{cal} < t_{tab} \rightarrow$ Accept H_0

$t_{cal} > t_{tab} \rightarrow$ Reject H_0

$4 > 2.95 \rightarrow$ Reject H_0

$\therefore$ Accept Alternative Hypothesis (H_1)

$\therefore$ There is significant difference b/w population & sample means.

---

11. The life a time of electric bulbs for a random sample of 10 from a large consignment given the following data.

Item	1	2	3	4	5	6	7	8	9	10
Life in hours	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

We can accept the Hypothesis that Average life time of is 4 hours test at 5% level of significance.

Sol :

Given that

Random sample of (n) = 10

Los = 5%

$t_{0.05}$

μ = 4000

Step 1 :

Null Hypothesis : There is no significant difference b/w average life time bulbs

$H_0 = \mu_1 = \mu_2$

Step 2 : There is significant difference b/w average lifetime of bulbs

$H_1 = \mu_1 \neq \mu_2$

Step 3 : Level of significance

= 5% Loss

$= t_{0.05/2} \Rightarrow t_{0.025}$

Degrees of freedom D.F = n – 1

= 10 – 1

$\boxed{V = 9}$

$t_{0.025} = 2.26$

Step 4 : Test statistic

$$t = \frac{|\bar{x} - \mu|\sqrt{n}}{s}$$

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} \quad n = 10; \ \mu = 4000$$

Item	x	$x-\bar{x}$	$(x-\bar{x})^2$
1	4.2	4.2 – 4.4 = – 0.2	0.04
2	4.6	4.6 – 4.4 = 0.2	0.04
3	3.9	3.9 – 4.4 = – 0.5	0.25
4	4.1	4.1 – 4.4 = – 0.3	0.09
5	5.2	5.2 – 4.4 = – 0.8	0.16
6	3.8	3.8 – 4.4 = – 0.6	0.36
7	3.9	3.9 – 4.4 = – 0.5	0.25
8	4.3	4.3 – 4.4 = – 0.1	0.01
9	4.4	4.4 – 4.4 = 0	0
10	5.6	5.6 – 4.4 = 1.2	1.44

$$\bar{X} = \frac{X}{N} = \frac{44}{10} = 4.4 \qquad \Sigma(X-\bar{X})^2 = 3.12$$

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$$s = \sqrt{\frac{3.12}{9}}$$

$$s = \sqrt{0.34}$$

$$s = 0.58$$

"s" value subtute in "t"

$$t = \frac{|x-\mu|\sqrt{n}}{s}$$

$$t = \frac{|44-4|\sqrt{10}}{0.58}$$

$$t = \frac{0.4\sqrt{10}}{0.58}$$

$$t = \frac{0.4(3.16)}{0.58}$$

$$t = \frac{1.26}{0.58}$$

$$\boxed{t_{cal} = 2.17}$$

Step 5 : Acceptence & Rejectence

$t_{cal} < t_{tab}$ Accept H_0

2.17 < 2.26 Accept H_0

$\therefore$ Accept null hypothesis

$\therefore$ There is no significant difference b/w Average life time of the bulbs.

Q24. Explain the procedure for testing of two mean.

Ans :

Given two independent random samples of size n_1 and n_2 with means $\bar{X}_1$ and $\bar{X}_2$ and standard deviations S_1 and S_2 we may be interested in testing the hypothesis that the samples come from the same normal population. To carry out the test, we calculate the statistic as follows :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

where $\bar{X}_1$ = mean of the first sample

$\bar{X}_2$ = mean of the second sample

n_1 = number of observations in the first sample

n_2 = number of observations in the second sample

S = combined standard deviation

The value of S is calculated by the following formula :

$$S = \sqrt{\frac{\Sigma\left(X_1 - \bar{X}_1\right)^2 + \Sigma(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

PROBLEMS

12. Two different types of drugs 'A' and 'B' were tried on certain patients for increasing weight. 6 persons were given drug 'A" and 8 persons were given drug 'B'. The increase in pounds is given below,

Drug 'A'	**7**	**10**	**13**	**12**	**4**	**8**		
Drug 'B'	**12**	**8**	**3**	**18**	**16**	**9**	**8**	**3**

Do the drugs 'A' and 'B' differ significantly with regard to their effect in increase in weight?

Sol :

Let the weights (in kgs) of the patients treated with drugs A and B be denoted by suitable variables X and Y respectively.

We set up the null hypothesis, $H_0 : \mu_x = \mu_y$ i.e., there is no significant difference between the drugs A and B, regards their effect on increase in patients weight.

Alternative hypothesis, $H_1 : u_x \neq \mu_y$

Under H_0, the appropriate test statistic is,

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Degree of freedom (d.f) = $t_{n_1 + n_2 - 2}$

Computation of Sample Means and Standard Deviations

Here, $n_1 = 6$, $\Sigma X = 54$, $\Sigma\left(X - \bar{x}\right)^2 = 56$, $\bar{x} = \frac{\Sigma x}{n_1} = \frac{54}{6} = 9$

$n_2 = 8$, $\Sigma y = 77$, $\Sigma\left(y - \bar{y}\right)^2 = 209.87$, $\bar{y} = \frac{\Sigma y}{n_2} = \frac{77}{8} = 9.625$

And, $S^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma\left(x - \bar{x}\right)^2 + \Sigma\left(y - \bar{y}\right)^2\right]$

$$= \frac{1}{6 + 8 - 2}[56 + 209.87]$$

$$= \frac{265.87}{12} = 22.16$$

$\therefore S = \sqrt{22.16} = 4.71$

$$\therefore t = \frac{\bar{x} - \bar{y}}{S\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{9 - 9.625}{4.71\sqrt{\left(\frac{1}{6} + \frac{1}{8}\right)}} = \frac{-0.625}{4.71\sqrt{\frac{7}{24}}} = \frac{-0.625}{2.54} = -0.25$$

Degree of freedom (d.f) = $t_{n_1 + n_2 - 2}$

$$= t_{6+8-2} = t_{12}$$

Hence, tabulated value of t for 12 d. f at 5% level of significance for the left tailed test is 1.782.

Thus, calculated $t = -0.25$, which is less than tabulated value of t, (i.e.,1.782).

Therefore, null hypothesis H_0 cannot be rejected at 5% level of significance and we may conclude that the drugs A and B do not differ significantly as regards their effect on increases in patients weights at 5% level of significance.

3.4.2 Paired t-test

Q25. Explain briefly about Paired t-test.

Ans : **(Jan.-20)**

Two samples are said to be dependent when the elements in one sample are related to those in the other in any significant or meaningful manner. In fact, the two samples may consist of pairs of observations made on the same object, individual or, more generally, on the same selected population elements. When samples are dependent they comprise the same number of elementary units. We may carry out some experiment, say, to find out the effect of training on some employees, find out the efficacy of a coaching class or determine whether there is a significant difference in the efficacy of two drugs - one made within the country and another imported. Often the use of dependent or paired samples will enable us to perform a more precise analysis, because they will allow us to control the entraneous factors. The t-test based on paired observations is defined by the following formula :

$$t = \frac{\bar{d} - 0}{S} \times \sqrt{n} \text{ or } t = \frac{\bar{d}\sqrt{n}}{S}$$

where $\bar{d}$ = the mean of the differences

S = the standard deviation of the differences

The value of S is calculated as follows :

$$S = \sqrt{\frac{\Sigma\left(d - \bar{d}\right)^2}{n-1}} \text{ or } \sqrt{\frac{\Sigma d^2 - n\left(\bar{d}\right)^2}{n-1}}$$

It should be noted that *t* is based on n - 1 degrees of freedom.

PROBLEMS

13. Eleven sales executive trainees are assigned selling jobs right after their recruitment. After a fortnight they are withdrawn from their field duties and given a month's training for executives sales. Sales executed by them in thousands of rupees before and after the training in the same period are listed below,

Sales (000 `) (Before training)	**23**	**20**	**19**	**21**	**18**	**20**	**18**	**17**	**23**	**16**	**19**
Sales (000 `) (After training)	**24**	**19**	**21**	**18**	**20**	**22**	**20**	**20**	**23**	**20**	**27**

Do these data indicate that the training has contributed to their performance ?

Sol :

Let the null hypothesis be that the executive trainees have not benefited by the training,

$H_0 : \mu_1 = \mu_2$

$H_0 : \mu_1 = \mu_2$

Alternate hypothesis training program has resulted in more sales.

Sample size small population variance not known

Paired t - test used,

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

$$s = \sqrt{\frac{\Sigma\left(d-\bar{d}\right)^2}{n-1}} = \sqrt{\frac{\Sigma d^2 - n\left(\bar{d}\right)^2}{n-1}}$$

$$= \sqrt{\left(\frac{\Sigma d^2}{n-1}\right) - \frac{\left(\Sigma d\right)^2}{n\left(n-1\right)}}$$

Sl.No.	Sales (1000) `		Difference	
	Before Training (1)	After Training (2)	d=(1) – (2)	d²
1	23	24	–1	1
2	20	19	1	1
3	19	21	–2	4
4	21	18	3	9
5	18	20	– 2	4
6	20	22	–2	4
7	18	20	– 2	4
8	17	20	– 3	9
9	23	23	0	0
10	16	20	– 4	16
11	19	27	–8	64
			Σd = – 20	Σd^2 = – 116

$$\text{Mean, } \bar{d} = \frac{\Sigma d}{n} = \frac{-20}{11} = -1.8181$$

$$S = \sqrt{\frac{\Sigma\left(d-\bar{d}\right)^2}{n-1}} = \sqrt{\frac{\Sigma d^2 - n\left(\bar{d}\right)^2}{n-1}}$$

$$= \sqrt{\frac{116-11\left(-1.8181\right)^2}{11-1}} = \sqrt{\frac{116-11\left(3.3055\right)}{10}}$$

$$= \sqrt{\frac{116-36.36}{10}} = \sqrt{\frac{79.64}{10}}$$

$= \sqrt{7.964}$

$= 2.822$

$\therefore S = 2.822$

$\therefore t = \dfrac{\bar{d}}{S/\sqrt{n}}$

$= \dfrac{-1.8181}{2.822/\sqrt{11}}$

$= \dfrac{-1.8181}{0.8508}$

$= -2.137$

Critical Value

Degree of freedom, (d.f) = n – 1

= 11 – 1

= 10

Decision

For degree of freedom (d.f) = 10 the table value of (5%)$t_{0.05}$ = 1.812. The table value is more than computed value hence null hypothesis is accepted that is training program has not contributed to sales.

14. To verify weather a course in accounting improved performance a similar test to 12 pastispents both before and after course the original marks recorded in alphabital order of partispents.

Before Couse	After Course
44	53
60	38
61	69
52	57
32	46
44	39
70	73
41	48
67	73
72	74
53	60
72	78

is the course is useful to the students ?

Sol :

Given

Total no. of students joined in course n = 12

Since n < 30 it is dependent variable "t" test

Step 1 :

Null Hypothesis : The course is not useful to the students

Step 2 :

Alternative Hypothesis : The course is useful to the students

Step 3 :

Level of significance

= 5% Loss

$= t_{0.05}$

Degrees of freedom D. F = n – 1 $\Rightarrow$ 11

$$\boxed{V = 11}$$

$$V = 11 \Rightarrow \frac{0.05}{2} = 0.025 \Rightarrow t_{0.025} = 2.201$$

Step 4 : Test statistic

Name of Students	Before Course (I)	After Course (II)	d= II - I	d^2
A	44	53	9	81
B	60	38	-22	484
C	61	69	8	64
D	52	57	5	25
E	32	46	14	196
F	44	39	–5	25
G	70	73	3	9
H	41	48	7	49
I	67	73	6	36
J	72	74	2	4
K	53	60	7	49
L	72	78	6	36
			40	1058

$$\bar{d} = \frac{\Sigma d}{n} = \frac{40}{12} \quad \Sigma d^2 = 1058$$

$$\bar{d} = 3.3$$

$$t_{cal} = \frac{\bar{d}.\sqrt{n}}{s}$$

$$S = \sqrt{\frac{\Sigma d^2 - n\left(\bar{d}\right)^2}{n-1}}$$

$$S = \sqrt{\frac{1058 - 12(3.3)^2}{12-1}}$$

$$S = \sqrt{\frac{1058 - 12(10.89)}{11}}$$

$$S = \sqrt{\frac{1058 - 130.68}{11}}$$

$$S = \sqrt{\frac{927.3}{11}}$$

s = 9.18

Substute "s" value in "t"

$$t_{cal} = \frac{\bar{d}.\sqrt{n}}{S}$$

$$t_{cal} = \frac{3.3\sqrt{12}}{s}$$

$$t_{cal} = \frac{(3.3)(3.4)}{9.18}$$

$$t_{cal} = \frac{11.2}{9.18}$$

$$\boxed{t_{cal} = 1.24}$$

Step 5 : Acceptence & Rejectence criteria

$t_{cal} < t_{tab}$

1.24 < 2.201

$\therefore$ Accept H_0

$\therefore$ The course is not useful to the students.

Exercises Problems

1. Draw a multiple bar diagram from the following data :

Year	Sales ('000 `)	Gross Profit ('000 `)	Net Profit ('000 `)
2005	120	40	20
2006	130	45	30
2007	140	55	35
2008	150	60	40

2. Represent the following data on details of cost of two commodities by the rectangular diagram :

Details	Commodity A	Commodity B
Price per unit	` 40	` 40
Quantity sold	40 units	30 units
Value of raw materials	` 520	` 500
Other expenses of production	` 640	` 600
Profits	` 440	` 400

8. The following table gives the results of 250 samples, all of same size, drawn from the same population for testing the null hypothesis H_0

	H_0 true	H_0 false	Total
Reject H_0	5	145	150
Accept H_0	90	10	100
Total	95	155	250

Find :

(i) Level of significance ;

(ii) Power of the test;

(iii) Critical region

[Ans : (i) $\frac{5}{250}$ = 0.02, (ii) 1 – $\frac{10}{250}$ = 0.96, (iii) 150 samples]

8. Random samples of 125 persons each have been drawn from two big cities A and B. The mean income and the standard deviation for the sample from city A are ` 2,000 and ` 100 respectively, for the sample from city B, the mean and standard deviation are ` 2,050 and ` 200 respectively.

Is there a significant difference in per capital income of the two cities (at 5% significance level) ? Assume the distribution of income to be normal for each city.

[Ans : $H_0 : \mu_1 = \mu_2$; $H_1 : \mu_1 \neq \mu_2$; |Z| = |–2.5|>1.96; Significant]

10. A consulting firm wants to decide at the 5% level of significance if the salaries of construction workers between Mumbai and Delhi random sample of 100 construction workers in Mumbai has an average weekly salary of ` 400 with a standard deviation of ` 100. In Delhi a random sample of 75 workers has an average weekly salary of ` 375 workers a standard deviation of ` 80. Is there a significant difference between salaries of construction workers in Mumbai and Delhi ?

[Ans : $H_0 : \mu_1 = \mu_2$; $H_1 : \mu_1 \neq \mu_2$; Z= 2.38, Significant at 5% level of significance (Not Significant at 1% level of significance]

Short Question and Answers

1. What is tabulation ? Explain the significance of tabulation.

Ans :

One of the simplest and most revealing devices for summarizing data and presenting them in a meaningful fashion is the statistical table . A table is a systematic arrangement of statistical data in columns and rows. Rows are horizontal arrangements whereas columns are vertical ones. The purpose of a table is to simplify the presentation and to facilitate ‘comparisons. The simplification results from the clear-cut and systematic arrangement, which enables the reader to quickly locate desired information.

The significance of tabulation will be clear from the following points:

(i) It simplifies complex data

When data are tabulated all unnecessary retails and repetitions are avoided. Data are presented systematically in columns and rows. Hence, the reader gets a very clear idea of what the table presents. There is thus a considerable saving in time taken in understanding what is represented by the data and all confusion is avoided. Also a large amount of space is saved because of non-duplicating of his headings and designations; the description at the top of a column serves for all the terms beneath it.

(ii) It facilitates comparison

Tabulation facilitates comparison. Since a cable is divided into various parts and for each part there are totals and sub-totals, the relationship between different parts of data can be studied much more easily with the help of a table than without it.

iii) It gives identity to the data

When the data are arranged in a table with a title and number they can be distinctly identified and can be used as a source reference in the interpretation of a problem.

iv) It reveals patterns

Tabulation reveals patterns within the figures which cannot be seen in the narrative form. It also facilitates the summation of the figures if the reader desires to check the totals.

2. Explain univariate tabulation with an example.

Ans :

In statistics tabulations can be either univariate, bivariate and multivariate in nature. Uni-variate analysis is one of the simplest form of statistical analyses which includes only one question. This one question has to be tabulated and this is called as a univariate tabulation. Univariate tables are usually used when the question itself gives the vital important for the analysis. Univariate table can be explained as follows,

Example

Find the percentage of the customers opinions regarding the product performance.

Response Category	Number of Responses	Percentage
Excellent	40	26.66
Good	50	33.33
Fair	30	20
Poor	30	20
Total	**150**	**100**

Sol :

33.33% of the responses opinion is good regarding the product and 20% opined to be fair and 20% opined for poor performance of the product.

3. Define classification.

Ans :

Classification is the grouping of related facts into classes. Facts in one class differ from those of classification. Sorting facts on one basis of classification and then on another basis is called cross-classification. This process can be repeated as many times as there are possible bases of classification. Classification of data is a function very similar to that of sorting letters in a post-office. It is well known that letters collected in a post-office are sorted into different lots on a geographical basis, i.e., in accordance with their destinations such as Mumbai, Calcutta, Kanpur, Jaipur, etc. They are then put into separate bags, each containing letters with a common characteristic, viz., having the same destination. Classification of statistical data is comparable to the sorting operation.

The principal objectives of classifying the data are :

➢ To condense the mass of data in such a manner that similarities and dissimilarities can be readily apprehended. Millions of figures can thus be arranged in a few classes having common features.

➢ To facilitate comparison.

➢ To pinpoint the most significant features of the data at a glance.

➢ To give prominence to the important information gathered while dropping out the unnecessary elements.

➢ To enable a statistical treatment of the material collected.

4. What are the difference between Classification and Tabulation?

Ans :

Classification	Tabulation
It is the basis for tabulation	It is the basis for further analysis
It is the basis for simplification	It is the basis for presentation
Data is divided into groups and subgroups on the basis of similarities and dissimilarities.	Data is listed according to a logical sequence of related characteristics
data are separated and grouped based on a property of the data common to all values	Data is arranged into columns and rows based on characteristics /properties, or indicators.
emphasize on the presentation aspects of the data	used as a means of sorting of data for further analysis

5. Define diagram and diagramatical represenation.

Ans :

(i) Diagrams

Diagram is a visual presentation of statistical information. The pictorial presentation helps in proper understanding the data . Diagrams are of different types like pie diagram, rectangles, lines, pictures and maps.

(ii) Diagrammatic Representation of Data

Apart from classification and tabulation methods of efficiently presenting the statistical data, the statistical data can also be presented in a convincing, appealing and easily understandable form using the diagrammatic representation (i.e., making use of diagrams).

6. Explain advantages and disadvantages of diagrams.

Ans :

Advantages

1. Diagrams give a clear picture of the data. Data presented with the help of diagrams can be understood and grasp even by a common man in a very short time.
2. Diagrammatical representation of data can be used universally at any place.
3. Diagrammatical representation not only saves time and energy but also is economical.
4. By using diagrammatic representation, data can be condensed.
5. By diagrammatic representation, comparison between data can be made without actually computing the statistical measures.
6. Diagrams are more impressive, attractive and fascinating compared to any other form of representation.
7. More information can be obtained using diagrammatic representation when compared to tabular representation of data.

8. Diagrams have an appealing effect to the user and are more easily remembered than any other representation.

Disadvantages

Some of the disadvantages of diagrammatic presentation are,

1. Diagrams do not show exact values.
2. Drawing diagrams is a difficult task.
3. Skills are required to present the data in a diagrammatic form.
4. Diagrammatic presentation of data reveals limited facts.
5. Diagrams are not substitute for tabular presentation.
6. Diagrams give only rough idea. Analyzing the data in detail is not possible through diagrams.
7. Diagrams drawn through false data give false idea.

7. Bar Diagram.

Ans :

These diagrams are very simple and easy to construct. They consider only the length of the bars but not their widths. These type of diagrams represents the data with the help of thick lines or bars.

Points to Remember and Rules to be Followed While Constructing Bar Diagrams

(i) Maintain uniformity in the width of the bars.

(ii) Represent the bars either vertically or horizontally. Easy comparison and better appearance is achieved with the help of vertical arrangement of bars,

(iii) Maintain uniformity in the gaps between the bars.

(iv) The bar diagrams must be drawn in such a way that every bar has its respective figure at the top. By doing so, we can avoid confusion among readers and make easy comparison of various objects.

Merits of Bar Diagram

Some of the merits of bar diagrams are as follows,

(i) Every individual irrespective of whether they are accustomed to reading charts or not, can understand bar diagrams.

(ii) Mass data can be compared effectively by using bar diagrams.

(iii) Compared to other diagrams, constructing bar diagrams is easy and simple.

8. applications of t-distribution.

Ans :

1. The variable t-distribution ranges from minus infinity to plus infinity.
2. The constant *c* is actually a function of *v* (pronounced as nu), so that for a particular value of *v*, the distribution of f(t) is completely specified. Thus f(t) is a family of functions, one for each value of *v*.
3. Like the standard normal distribution, the t-distribution is symmetrical and has a mean zero.
4. The variance of the t-distribution is greater than one, but approaches one as the number of degrees of freedom and, therefore, the sample size becomes large. Thus the variance of the t-distribution approaches the variance of the standard normal distribution as the sample size increases. It can be demonstrated that from an infinite number of degrees of freedom ($v = \infty$), the t-distribution and normal distribution are exactly equal.

9. Applications of t-distribution.

Ans :

The following are some important applications of t-distribution,

(i) Test of hypothesis about the population mean.

(ii) Test of hypothesis about the difference between two means.

(iii) Test of hypothesis about the difference between two means with dependent samples.

(iv) Test of hypothesis about coefficient of correlation.

10. Explain briefly about Paired t-test.

Ans :

Two samples are said to be dependent when the elements in one sample are related to those in the other in any significant or meaningful manner. In fact, the two samples may consist of pairs of observations made on the same object, individual or, more generally, on the same selected population elements. When samples are dependent they comprise the same number of elementary units. We may carry out some experiment, say, to find out the effect of training on some employees, find out the efficacy of a coaching class or determine whether there is a significant difference in the efficacy of two drugs - one made within the country and another imported. Often the use of dependent or paired samples will enable us to perform a more precise analysis, because they will allow us to control the entraneous factors. The t-test based on paired observations is defined by the following formula :

$$t = \frac{\bar{d} - 0}{S} \times \sqrt{n} \text{ or } t = \frac{\bar{d}\sqrt{n}}{S}$$

where $\bar{d}$ = the mean of the differences

S = the standard deviation of the differences

The value of S is calculated as follows :

$$S = \sqrt{\frac{\Sigma\left(d - \bar{d}\right)^2}{n-1}} \quad \text{or} \quad \sqrt{\frac{\Sigma d^2 - n\left(\bar{d}\right)^2}{n-1}}$$

It should be noted that *t* is based on n - 1 degrees of freedom.

11. What is tabulation ? Explain the signifi- cance of tabulation.

Ans :

Differentiate between diagrams and graphs.

Ans :

The differences between Diagrams and Graphs are as follows,

Diagrams	Graphs
1. Diagrams are usually constructed on a plain sheet and they are useful for comparison purpose.	1. Graphs are constructed on a special sheet known as graph paper and they are used for studying the relationship between the variables
2. In Diagrammatic presentation, data is presented by using bars, pie charts, circles, cubes etc.	2. In Graphical presentation, data is presented through lines or points of various types like dots, dashes, dot-dash and so on.
3. Diagrams give aproximate idea and are less accurate.	3. Graphs provide reliable, accurate and authentic information.
4. Compared to graphs, constructing diagrams is a difficult task.	4. They are easy to construct compared to diagrams.
5. Geographical and categorical data can be presented through diagrams.	5. Data concerning time series and frequency distribution can be presented through graphs.

UNIT IV	(a) Analysis of Variance - One Way and Two-Way ANOVA (with and without Interaction). Chi-Square distribution: Test for a specified Population variance, Test for Goodness of fit, Test for Independence of Attributes. (b) Correlation Analysis- correlation, limits for coefficient of Correlation, Karl Pearson's coefficient of correlation, Spearman's Rank correlation, Linear and Multiple regression analysis, Discriminant analysis, Exploratory Factor Analysis. Introduction to Structural Equation Modeling, Cluster Analysis and Conjoint Analysis.

4.1 Analysis of Variance

Q1. What is ANOVA? What are its assumptions and applications?

Ans : **(Jan.-20, Imp.)**

ANOVA

The variance test is also known as ANOVA. ANOVA is the acronym for Analysis of Variance. Analysis of variance is a statistical technique specially designed to test whether the means of more than two quantitative populations are equal i.e., to make inferences about whether those samples are drawn from the populations having the same mean.

The test is called 'F' test as it was developed by R.A Fisher in 1920's. The test is conducted in situations where we have three or more to consider, at a time an alternative procedure (to t-test) needed for testing the hypothesis that all samples could likely be drawn from the same population.

Example

Five fertilizers are applied to four plots, each of wheat and yield of wheat on these plots is given. We are interested in finding out whether the effects of these fertilizers on the yields are significantly different or, in other words, whether the samples have come from the same population. ANOVA answers this question.

Assumptions for ANOVA

Analysis of variance test is based on the test statistic F (or variance ratio).

It is based on the following assumptions,

(i) Observations are independent.

(ii) Each sample is drawn randomly from a normal population as the sample statistics reflect the characteristic of the population.

(iii) Variance and means are identical for those population from which samples have been drawn.

Applications of ANOVA

The applications of ANOVA are as follows,

1. Anova is used in education, industry, business, psychology fields mainly in their experiment design.
2. Anova helps to save time and money as several population means can be compared simultaneously.
3. Anova is used to test the linearity of the fitted regression line and correlation ratio, significance test statistic of anova

$$= F(r-1, n-r).$$

4.1.1 One Way and Two-Way ANOVA (with and without Interaction)

Q2. Explain briefly about One Way ANOVA.

Ans : **(Jan.-20, Imp.)**

ANOVA One Way Classification

In these classification the data is classified according to only one criteria i.e., It includes only one factor.

Steps involved in ANOVA One Way

(i) Null Hypothesis $H_0 = \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$ (where means are equal)

(ii) Alternative Hypothesis $H_1 = \mu_1 \neq \mu_2 ... \neq \mu_k$ (when means are not equal)

Arithmetic mean and drawn from the means of population from which "K" samples are drawn which are equal to another.

Step 1

(i) Calculation variance between samples

(ii) Calculation of grand average $\bar{\bar{X}}$

$$\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{N}$$

(iii) Take the difference between means of variance samples of grand average.

(iv) Square the deviations and obtain total which will give some of squares between samples.

(v) Divide the total by degrees of freedom K = No. of samples $\boxed{V = K - 1}$

Step 2

(i) Calculate variance within samaples.

(ii) Calculation of mean value of samples $\bar{X}_1, \bar{X}_2, ...$

(iii) Take the deviation of variance items in α samples from mean values.

(iv) Square the derivations and obtain total which give sum of squares within samples.

(v) Divide the total by degrees of freedom

$$\boxed{V = N - K}$$

N = No. of observations

K = Refers to the no. of samples

Step 3

Calculation of "F" Ratio

$$\text{“F”} = \frac{S_1^2}{S_2^2}$$

Step 4

Compare the calculated value of "F" with 5% level of significance.

- If $F_{cal} > F_{tab} \rightarrow$ Difference between sample means are significant.
- If $F_{cal} < F_{tab} \rightarrow$ Difference between sample means are not significant.

Step 5

Sources of Variations	SS (Sum of Squares)	V = Degress of Freedom	MS Mean Squares	"F" Ration
Between Samples	SSC	$V_1 = C - 1$	$MSC = \frac{SSC}{C-1}$	$\frac{MSC}{MSE}$
Within Samples	SSE	$V_2 = M - C$	$MSE = \frac{SSE}{M-C}$	
Total	SST	n – 1		

SSC = Sum of Squares between samples (columns)

SSE = Sum of Squares within samples (rows)

SST = Total Sum & Squares of Variations

MSC = Mean sum of squares between samples

MSE = Mean sum of squares within samples.

PROBLEMS

1. Test whether the significance of possible variation in performance in a certain test between the grammer schools of a City A common test is given to number of students taken of random from the senior Vth Class of four schools.

	Schools			
	A	**B**	**C**	**D**
1	**8**	**12**	**18**	**13**
2	**10**	**11**	**12**	**9**
3	**12**	**9**	**16**	**12**
4	**8**	**14**	**6**	**16**
5	**7**	**4**	**8**	**15**

Sol :

Given

No. of samples (k) = 4 samples (A, B, C, D)

Hypothesis

Null Hypothesis : There is no significant difference between schools

$H_0 = \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

Alternative Hypothesis : There is significant difference between schools

$H_1 = \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

Step 1: Calculation of Variance between Samples

Calculation of $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$, $\bar{X}_4$

X_1	X_2	X_3	X_4
8	12	18	13
10	11	12	9
12	9	16	12
8	14	6	16
7	4	8	15
$\bar{X}_1 = 9$	$\bar{X}_2 = 10$	$\bar{X}_3 = 12$	$\bar{X}_4 = 13$

Calculation of grand mean $\bar{\bar{X}}$

$$\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \bar{X}_4}{N}$$

$$\bar{\bar{X}} = \frac{9+10+12+13}{4}$$

$$\boxed{\bar{\bar{X}} = 11}$$

Calculation of variance between samples

Sample (A) $(\bar{X}_1 - \bar{\bar{X}})^2$	**Sample (B)** $(\bar{X}_2 - \bar{\bar{X}})^2$	**Sample (C)** $(\bar{X}_3 - \bar{\bar{X}})^2$	**Sample (D)** $(\bar{X}_4 - \bar{\bar{X}})^2$
$(9-11)^2=4$	$(10-11)^2=1$	$(12-11)^2=1$	$(13-11)^2=4$
$(9-11)^2=4$	$(10-11)^2=1$	$(12-11)^2=1$	$(13-11)^2=4$
$(9-11)^2=4$	$(10-11)^2=1$	$(12-11)^2=1$	$(13-11)^2=4$
$(9-11)^2=4$	$(10-11)^2=1$	$(12-11)^2=1$	$(13-11)^2=4$
$(9-11)^2=4$	$(10-11)^2=1$	$(12-11)^2=1$	$(13-11)^2=4$
20	5	5	5

Sum of the squares of between samples

= 20 + 5 + 5 + 20

= 50

Mean sum of the squares of between samples

$$= \frac{50}{K-1}$$

$$= \frac{50}{4-1}$$

$$= \frac{50}{3}$$

= 16.6

Step 2 : Calculation Variance with in Samples

Sample (A)		Sample (B)		Sample (C)		Sample (D)	
X_1	$(X_1-\bar{X}_1)^2$	X_2	$(X_2-\bar{X}_2)^2$	X_3	$(X_3-\bar{X}_3)^2$	X_4	$(X_4-\bar{X}_4)^2$
8	1	12	4	18	36	13	0
10	1	11	1	12	0	9	16
12	9	9	1	16	16	12	1
8	1	14	16	6	36	16	9
7	4	4	36	8	16	15	4
	16		58		104		30

Sum of the square within samples

$= 16 + 58 + 104 + 30$

$= 208$

Mean sum of the squres of within samples

$$= \frac{208}{M-K}$$

M = Total no. of observations

K = No. of samples

$$= \frac{208}{20-4} = \frac{208}{16} = 13$$

Step 3 : Calculation of "F" Ratio

Sources of Variations	Sum of Squares	Degress of Freedom	MS	"F" Ration
Between Samples	50	$V_1 = C-1$ $V_1 = 4-1$ $V_1 = 3$	$MSC = \frac{SSC}{C-1}$ 16.6	$F = \frac{MSC}{MSE}$ $F = 1.27$
Within Samples	208	$V_2 = M-C$ $V_2 = 20-4$ $V_2 = 16$	$MSE = \frac{SSC}{M-C}$ 13	
Total	258	19		

Step 4 : Acceptance & Reject

(V_1 = 3, V_2 = 16 $t_{0.05}$ = 3.24)

$F_{cal} < F_{tab}$

1.24 < 3.24

∴ Accept H_0

∴ **There is no significance difference between schools.**

2. Four machines A, B, C and D are used to produce a certain kind of cotton fabrics. Samples of size 4 with each unit as 100 square metres are selected from the outputs of the machines at random and the number of flaws in each 100 square metres show the following result.

A	B	C	D
8	**6**	**14**	**20**
9	**8**	**12**	**22**
11	**10**	**18**	**25**
12	**4**	**9**	**23**

Do you think that there is a significant difference in the performance of the four machines?

Sol :

Let the null hypothesis be that there is no significant difference in the performance of the four machines i.e., $\mu_0 = \mu_1 = \mu_{11} = \mu_{111}$.

A	B	C	D	Total
8	6	14	20	48
9	8	12	22	51
11	10	18	25	64
1 2	4	9	23	48
40	28	53	90	GT = 211

Correction Factor (C.F) = $\frac{(GT)^2}{N}$ = N = 16

C.F. = $\frac{(211)^2}{16}$ = 2782.56

Total Sum of Squares (TSS)

$= \Sigma_i \Sigma_j X_{ij}^2 - C.F$

$= [(8)^2 + (6)^2 + (14)^2 + (20)^2 + (9)^2 + (8)^2 + (12)^2 + (22)^2 + (11)^2 + (10)^2 + (18)^2.$

$+ (25)^2 + (12)^2 + (4)^2 + (9)^2 + (23)^2] - 2782.56$

$= [64 + 36 + 196 + 400 + 81 + 64 + 144 + 484 + 121 + 100 + 324 + 625 + 144 + 16 + 81 + 529] - 2782.56$

$= 3409 - 2782.56 = 626.44$

Sum of squares between the samples,

$$(SSB) = \sum_j \frac{T_j^2}{n_j} - \frac{(GT)^2}{N}$$

$$= \left[\frac{(40)^2}{4} + \frac{(28)^2}{4} + \frac{(53)^2}{4} + \frac{(90)^2}{4}\right] - 2782.56$$

$= [400 + 196 + 702.25 + 2025] - 2782.56$

$= 3323.25 - 2782.56$

$= 540.69$

Sum of squares within samples,

SSW $= TSS - SSB$

$= 626.44 - 540.69$

$= 85.75$

ANOVA Table

Sources of Variation	Degree of Freedom	Sum of Squares	Mean Square	F-Ratio
Between Samples	$(k-1) = (4-1) = 3$	540.69	$\frac{540.69}{3} = 180.23$	$\frac{180.23}{7.14} = 25.22$
Within samples	$(n-k) = 16-4 = 12$	85.75	$\frac{85.75}{12} = 7.14$	
Total	$(n-1) = 16-1 = 15$			

F-ratio$_{(3, 12)}$ = calculated = 25.22

F-ratio from table $V_1 = 3$ and $V_2 = 12$ at 5% level of significance = 3.49

Since $F_{(3, 12)}$ calculated > $F_{(3, 12)}$ table value we reject H_0 which means that there is a significant difference between the performance of four machines.

Q3. Explain briefly about two way ANOVA with and without interation?

Ans : **(Jan.-20)**

Two way classification/two factor ANOVA is defined where two independent factors have an effect on the response variable of interest.

Example : Yield of crop affected by type of seed as well as type of fertilizer.

Procedure

(a) Calculate the variance between columns,

$$SSC = \sum_{j=1}^{c} \frac{T_j^2}{n_j} - \frac{T^2}{N}$$

(b) Calculate the variance between rows,

$$SSR = \sum_{i=1}^{r} \frac{T_i^2}{n_i} - \frac{T^2}{N}$$

(c) Compute the total variance,

$$SST = \sum X_{ij}^2 - \frac{T^2}{N}$$

(d) Calculate the variance of residual or error,

$$SSE = TSS - (SSC + SSR)$$

(e) Divide the variances of between columns, between rows and residue by their respective degrees of freedom to get the mean squares.

(f) Compute F ratio as follows,

F-ratio concerning variation between columns,

$$= \frac{\text{Mean square between columns}}{\text{Mean squares of residual}}$$

F-ratio concerning variation between rows,

$$= \frac{\text{Mean square between rows}}{\text{Mean squares of residual}}$$

(g) Compare F-ratio calculated with F-ratio from table,

If F-ratio (calculated) < F-ratio (table), H_1 accepted,

If F-ratio (calculated) $\geq$ F-ratio (table), H_0 rejected,

H_1 accepted $\Rightarrow$ no significant differences

H_0 rejected $\Rightarrow$ significant differences

Two-Way ANOVA with Interation

Under two-way ANOVA with interaction, the total sum of squares SST is divided into four components, which are as follows,

1. The Sum of Squares for factor 'A' (SSA)
2. The Sum of Squares for factor 'B' (SSB)
3. The Sum of Squares for the interaction between two factors 'SSAB'.
4. The Error of Sum of Squares (SSE).

These factors can be represented as,

$$SST = SSA + SSB + SSAB + SSE$$

The main purpose of using two-way ANOVA with interaction is to understand the relationship between factors 'A' and 'B'. Such relationship will help to find out the impact, effect or influence of factor 'A' on factor 'B' and factor 'B' on factor 'A'.

Two-Way ANOVA without Interaction

Under two-way ANOVA without interaction, the total variability of data is divided into three components, which are as follows,

1. Treatment i.e., factor 'A'
2. Block i.e., factor 'B'
3. Chance.

However, the term 'block' refers to a matched group of observations from each population. When units of each block are assigned randomly to each treatment then the design of such experiment is referred as randomized block design.

Note :

Two factors are said to interact if the difference between levels (treatments) of one factor depends on the level of the other factor. Factors that do not interact are called additive.

A combination of a treatment from one factor with a treatment from another factor results in an interaction.

An interaction between two factors exists when for atleast one combination of treatments, the effect of combination is not additive.

ANOVA Table for Two-way Clasisifed Data with m-Observation Per Cell

Sources of Variation	Degree of Freedom	S.S	M.S.S	Variance Ratio F
Factor A	p – 1	S_A^2	$S_A^2 = \frac{S_A^2}{p-1}$	$F_A = \frac{S_A^2}{S_E^2}$
Factor B	q – 1	S_B^2	$S_B^2 = \frac{S_B^2}{q-1}$	$F_B = \frac{S_B^2}{S_E^2}$
Interaction AB	(p–1) (q–1)	S_{AB}^2	$F_{AB} = \frac{S_{AB}^2}{S_E^2}$	
Factor AB	pq (m – 1)	S_E^2	$S_E^2 = \frac{S_E^2}{pq(m-1)}$	
Total	pqm – 1			

Remark

The calculation of various sum of squares is facilitiated to a great extent by the use of following formulae,

$$C.F = \frac{G^2}{pqm} = \frac{T^2}{pqm}$$

$$TSS = \sum_i \sum_j \sum_k x_{ij}k^2 - C.F = RSS - CF = \sum_i T_i^2$$

$$S_A^2 = \frac{i}{M} CF$$

$$S_B^2 = \frac{j}{M} CF$$

$$S_{AB}^2 = \sum \sum T_{ij}^{\ 2}$$

SS due to Means (SSM)

$$= \frac{ij}{mp} CF$$

$$S_{AB}^2 = SSM - S_A^2 - S_B^2$$

$$S_E^2 = TSS - S_A^2 - S_B^2 - S_{AB}^2$$

Hypothesis Tests in Two-way ANOVA

- **Factor A Test :** Hypothesis is designed to determine whether there are any factor A main effects. Null Hypothesis true if and only if there are no differences in means due to different treatments (population) of factor A.
- **Factor B Test :** Hypothesis is test designed to detect factor B main effects. Null hypothesis is true if and only if there are no differences in means due to different treatments (populations) of factor B.
- **Test for AB Interactions :** Test for existence of interactions between levels of the two factors Null hypothesis is true if and only if there are no two way interactions between levels of factor A and levels of factor B, means factor effects are additive for two way ANOVA.

PROBLEMS

Two Way ANOVA

3. Four different drugs have been developed for a certain disease. These drugs are used under three different environments. It is assumed that the environment might affect efficacy of drugs. The number of cases of recovery from the disease per 100 people who have taken the drugs is tabulated as follows :

Environment	Drug A1	Drug A2	Drug A3	Drug A4
I	19	8	23	8
II	10	9	12	6
III	11	10	13	16

Test whether the drugs differ in their efficacy to treat the disease, also whether there is any effect of environment on the efficacy of disease.

Sol :

Null Hypothesis

H_o = There is no significant difference in the efficacy of drugs to treat the disease.

H_o = There is no significant effect of environment on the efficacy of disease.

Environment	Drug				Total
	A_1	A_2	A_3	A_4	
I	19	8	23	8	58
II	10	9	12	6	37
III	11	10	13	16	50
Total	**40**	**27**	**48**	**30**	**GT=145**

$$\text{Correction Factor, (CF)} = \frac{(\text{Grand Total})^2}{N}$$

$$= \frac{(145)^2}{12} = 1752.08$$

Total Sum of Squares (TSS)

$$(\text{TSS}) = \sum_i \sum_j X_{ij}^2 - \text{C.F}$$

$$= [(19)^2 + (8)^2 + (23)^2 + (8)^2 + (10)^2 + (9)^2 + (12)^2 + (6)^2 + (11)^2 + (10)^2 + (13)^2 + (16)^2] - \text{C.F}$$

$$= 2025 - 1752.08$$

$$\therefore \quad \text{TSS} = 272.92$$

Sum of Squares Between Drugs (Column)

$$\text{SSC} = \sum_j \frac{T_j^2}{n_j} - \frac{(GT)^2}{N}$$

$$= \left[\frac{(40)^2}{3} + \frac{(27)^2}{3} + \frac{(48)^2}{3} + \frac{(30)^2}{3}\right] - 1752.08$$

$$= (533.33 + 243 + 768 + 300) - 1752.08$$

$$= 1844.33 - 1752.08$$

$$\therefore \quad \text{SSC} = 92.25$$

Degree of freedom (r) = (C – 1)

= (4 – 1)

= 3

Sum of Squares Between Environment (Rows)

$$SSR = \sum_i \frac{T_i^2}{n_i} = \frac{(GT)^2}{N}$$

$$= \left[\frac{(58)^2}{4} + \frac{(37)^2}{4} + \frac{(50)^2}{4}\right] - C.F$$

$$= (841 + 342.25 + 625) - 1752.08$$

$$= 1808.25 - 1752.08$$

$\therefore$ SSR = 56.17

Degree of freedom,

$V_m = (r - 1) - (3 - 1) - 2$

Residual = Total sum of squares – (Sum of squares between columns + Sum of squares between rows)

= TSS – (SSC + SSR) = 272.92 – (92.25 + 56.17)

= 272.92 – 148.42 = 124.5

ANOVA TABLE

Sources of variation	Sum of squares	Degrees of Freedom	Means quares	Variance Ratio (F)
Between Drugs	92.25	(C–1) = (4–1) = 3	$\frac{92.25}{3} = 30.75$	$F_{(3.6)} = \frac{30.75}{20.75} = 1.48$
Between Environment	56.17	(r–1) = (3–1) = 2	$\frac{56.17}{2} = 28.09$	$F_{(2,6)} = \frac{28.09}{20.75} = 1.354$
Residual or Error (e)	124.5	(C–1) (r–1) = 3 × 2 = 6	$\frac{124.5}{6} = 20.75$	
Total	**272.92**	**(12 – 1) = 11**		

[**Note:** As level of significance is not given in the problem, assume 5% level of significance]

Critical value of $F_{0.05}$	Computed value of F
Drugs at V_0 (3,6) = 4.76	1.48
Environment at V_m(2,6) = 5.14	1.354

Table values are calculated as per 5% level of significance.

Decision

1. Drugs

Since the calculated value of F(1.48) is less than the table value (4.76), null hypothesis is accepted. Hence, there is no significant difference in the efficacy drugs.

2. Environment

Since the calculated value of F(1.354) is less than the table Value (5.14), null hypothesis is accepted. Hence, there is no affect of environment on the efficacy of disease.

4. Suppose that we are interested in establishing the yield producing ability of four types of soya beans A, B, C and D. We have three blocks of land X,Y and Z which may be different in fertility. Each block of land is divided into four plots and the different types of soya beans are assigned to the plots in each block by a random procedure. The following results are obtained:

	Soya Bean			
Block	**Type A**	**Type B**	**Type C**	**Type D**
X	5	9	11	10
Y	4	7	8	10
Z	3	5	8	9

Test whether A,B,C and D are significantly different.

Sol :

Null Hypothesis

H_o : There is no significant difference between A,B,C and D.

	Soya bean				
Block	**Type A**	**Type B**	**Type C**	**Type D**	**Total**
X	5	9	11	10	35
Y	4	7	8	10	29
Z	3	5	8	9	25
Total	12	21	27	29	GT = 89

$$\text{Correction Factor (CF)} = \frac{(\text{Grand Total})^2}{N} = \frac{(89)^2}{12}$$

$$= 660.08$$

Total Sum of Squares (TSS)

$$= \sum_i \sum_j X_{ij}^2 - \frac{(GT)^2}{N}$$

$$= [(5)^2 + (9)^2 + (11)^2 + (10)^2 + (4)^2 + (7)^2 + (8)^2 + (10)^2 + (3)^2 + (5)^2 + (8)^2 + (9)^2 1] - 660.08$$

$$= [25 + 81 + 121 + 100 + 16 + 49 + 64 + 100 + 9 + 25 + 64 + 81] - 660.8$$

$$= 735 - 660.08$$

$$\therefore \text{TSS} = 74.92$$

Sum of Squares Between Soya Bean (Columns)

$$SSB = \sum_j \frac{T_j^2}{n_j} - \frac{(GT)^2}{N}$$

$$= \frac{(12)^2}{3} + \frac{(21)^2}{3} + \frac{(27)^2}{3} + \frac{(29)^2}{3} - 660.08$$

$$= [48 + 147 + 243 + 280.33] - 660.08$$

$$= 718.33 - 660.08$$

$\therefore$ SSB = 58.25

Degree of freedom (r) = (K – 1)

= (4 – 1)

= 3

Mean sum of squares between the soya beans $= \frac{58.25}{3} = 19.42$

Sum of Squares within Blocks (SSW)

SSW = TSS – SSB

= 74.92 – 58.25

= 16.67

Mean sum of squares within the blocks

$$= \frac{16.67}{12-4} = \frac{16.67}{8} = 2.08$$

ANOVA TABLE

Sources of variation	Sum of squares	Degrees of Freedom	Means squares
Between soya bean type	58.25	(k – 1) = (4 – 1) = 3	19.42
Within blocks	16.67	(n – k) = (12 – 4) = 8	2.08
Total		(n – 1) = (12 – 1) = 11	

$$\text{F-Ratio} = \frac{\text{Mean square between soya bean type}}{\text{Mean square within blocks}}$$

$$= \frac{19.42}{2.08} = 9.34$$

[Note: Assuming level of significance as 5%]

$\text{F-Ratio}_{(3, 8)}$, calculated = 9.34

F-Ratio from table for V_1 = 3 and V_2 = 8 at 5% level of significance = 4.07

Decision

The calculated value of F is more than the table value. Therefore we reject null hypothesis (H_0) which means that there is a significant difference between types of soya beans.

5. A tea company appoints four salesman A, B, C, D and observes their sales in three seasons summer, winter and Monsoon. The figures (in lakhs of rupees) are given in the following table. Explain,

(a) Do the salesmen differ significantly in performance

(b) Is there significant difference between the seasons

Seasons	**Salesmen**				
	A	**B**	**C**	**D**	**Total**
Summer	36	36	21	36	129
Winter	28	29	31	31	119
Monsoon	26	28	29	29	112
Total	90	93	81	96	360

Support your answer with appropriate statistical analysis.

Sol :

By coding the data, we can simplify the task. Let us substract 30, based on the average of all the values given in problem, from all the observations. We get,

	Decoded data (X_{ij} – 30)				**Total**
	A	**B**	**C**	**D**	
Summer	6	6	–9	6	9
Winter	–2	–1	1	1	–1
Mansoon	–4	–2	–1	–1	–8
Total	0	3	–9	6	GT = 0

$$\text{Correction factor (C.F)} = \frac{(GT)^2}{N}$$

$$= \frac{(0)^2}{12} = 0$$

Total sum of squares,

$$TSS = \sum_j \sum_j X_{ij}^2 - C.F$$

$$= [(6)^2 + (6)^2 + (-9)^2 + (6)^2 + (-2)^2 + (-1)^2 + (1)^2 + (1)^2 + (-4)^2 + (-2)^2 + (-1)^2 + (-1)^2 - 0$$

$$= 36 + 36 + 81 + 36 + 4 + 1 + 1 + 1 + 16 + 4 + 1 + 1 - 0$$

$$= 218 - 0 = 218$$

Sum of Squares between Salesman (Columns)

$$SSC = \sum_j \frac{T_j^2}{n_j} - \frac{(GT)^2}{N}$$

$$= \frac{(0)^3}{3} + \frac{(3)^2}{3} + \frac{(-9)^2}{3} + \frac{(6)^2}{3} = 0$$

$$= 0 + 3 + 27 + 12 - 0 = 42$$

$\therefore$ SSC = 42

Degree of freedom (r) = (c – 1) = (4 – 1) = 3

Sum of Squares between Seasons (Rows)

$$SSR = \sum_j \frac{T_i^2}{n_i} - \frac{(GT)^2}{N} = \frac{(9)^2}{4} + \frac{(-1)^2}{4} + \frac{(-8)^2}{4} - 0$$

$$= 20.25 + 0.25 + 16 - 0 = 36.5$$

$\therefore$ SSR = 36.5

Degrees of freedom (C) = (r – 1) = (3 – 1) = 2

Residual = Total sum of square – (Sum of squares between columns + Sum of squares between rows)

= TSS – (SSC + SSR)

= 218 – (42 + 36.5) = 139.5

Anova Table

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	Variation Ratio or F = S_m^2 / S_e^2
Between sales men	42	3	$\frac{42}{3} = 14 = (S_0^2)$	$\frac{14}{23.25} = 0.602$
Between seasons	36.5	2	$\frac{36.5}{2} = 18.25 = (S_m^2)$	$\frac{18.25}{23.25} = 0.785$
Residual or error	139.5	(C–1) × (r–1) 3 × 2 = 6	$\frac{139.5}{6} = 23.25 - (S_e^2)$	
Total	**218**	**11**		

Critical Value of $F_{0.05}$	Computed Value of F
Salesmen at V_0 (3, 6) = 4.76	0.602
Salesmen at V_m (2, 6) = 5.14	0.784

Decisions

(a) **Salesmen :** Since the calculated value of F(0.602) is less than the tabulated value of F(4.76), null hypothesis (H_O) is accepted which states that there is no significant difference in the performance of salesmen.

(b) **Seasons :** Since the calculated value of F(0.784) is less than the tabulated value of F(5.14), null hypothesis (H_O) is accepted. According to this there is no significant difference between the seasons.

4.2 Chi-Square Distribution

Q4. Explain briefly about Chi-Square Distribution.

Ans : **(Imp.)**

The square of a standard normal variable is called a chi-square (χ^2) variate with one degree of freedom (d.f)

$$\chi^2 = \left(\frac{x-\mu}{\sigma}\right)^2$$

Where 'χ' is a random variable following normal distribution with mean 'μ' is an standard deviation σ. If χ_1, χ_2, ... , χ_v are independent random variables following normal distribution with means μ_1, μ_2, ... μ_v and standard dartrous σ_1, σ_2, ... σ_v then,

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

χ^2 is the sum of square of V independent standard normal variates following chi-square distribution. χ^2 is a family of distributions one for each value of V.

Applications of χ^2-Distribution

Chi-square distribution has a number of applications,

1. Chi-square test of goodness of fit.
2. Chi-square test for independence of attributes.
3. To test if population has a specified value of variance.
4. To test equality of several population proportions.

Chi-square Test Statistic

Chi-square is an important non-parametric test and does not have any prerequisites (no assumptions) in respect of the type of populations.

χ^2 describes the magnitude of discrepancy between theory and observations, Whether it can be attributed to chance or it is due to inadequacy of theory to fit the observed facts. If χ^2 is zero, observed and expected frequencies are equal. The greater χ^2 value more would be the discrepancy between observed and expected frequencies.

$$\chi^2 = S\frac{(O-E)^2}{e}$$

Where ,

O = Observed frequency

E = Expected frequency

Calculated value of χ^2 is compared with table value of χ^2 for given degrees of freedom at a specified level of significance.

(a) If χ^2 (calculated value) > χ^2 (table value) the difference between theory and observation is considered to be significant (fluctuations of sampling not responsible).

(b) If χ^2 (calculated value) < χ^2 (table value) the difference between theory and observation is not considered significant (fluctuations in sampling responsible)

Degrees of freedom have restriction of the total sample observations.

Assumptions/Conditions of χ^2 Test

There are five conditions to fulfill for a chi-square test,

1. Sample observation data must be independent of each other.
2. Random sampling from specified population to give sample data.
3. Data should not be in percentage or ratio form but original units to make comparison easy.
4. Sample size should have atleast 50 observations.

Q5. Explain the uses of Chi-square distribution.

Ans :

Use of Chi-Square Test

1. A. chi-square statistic can be used to test research questions involving cross-tabulated categorical variables.
2. An overall chi-square statistic is computed by summing the individual cell values (chi-squares) in a cross-tabulated table.
3. The degrees of freedom for a cross-tabulated table are row minus one times column minus one, i.e., df = (r – 1) (c – 1).
4. The chi-square test of independence can be used for any number of rows and columns, as long as the expected cell frequency is greater than five.
5. A chi-square test of independence is used to determine whether or not the rows and columns are independent (null hypothesis).
6. If the null hypothesis is true, it is still possible that the chi-square test could lead to a rejection of the null hypothesis (Type I error).
7. If the null hypothesis is false, it is still possible that the chi-square test could lead to retaining the null hypothesis (Type II error).
8. The ratio of each cell value to the overall chi-square value provides a variance accounted for interpretation of how much each cell contributed to the overall chi-square value.

4.2.1 Test for a Specified Population Variance

Q6. Explain briefly about Test for a Specified Population Variance.

Ans :

We want to test if the given normal population has specified variance $\sigma^2 = \sigma_0^2$ (say).

We set up Null Hypothesis $(H_0): \sigma^2 = \sigma_0^2$.

Alternative Hypothesis $(H_1): \sigma^2 \neq \sigma_0^2$.

Test Statistic

$$\chi^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sigma_0^2} = \frac{ns^2}{\sigma_0^2} \sim$$ a Chi Square Distri-bution with (n–1) D.F. The value of $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ is the sample variance.

Conclusion

By comparing the calculated value of Chi-Square with the tabulated value of Chi-Square for (n–1) D.F. at the selected Level of Significance, we may accept or reject the H_0.

Problems

6. A manufacturer's representative claims that his company has perfected an adjustment to a machine that will reduce the variability in the diameter of the screws produced by the machine. The original variance was 0.010 inches. To determine whether or not his claim is reasonable, a random sample of 20 screws is selected from those produced after the adjustment has been made. The sample variance is 0.009. Does this evidence support the representative's claim ?

Sol. :

1. Null Hypothesis (H_0) : s^2 = 0.010

   Alternative Hypothesis (H_1) : $s^2 \neq 0.010$

2. The data given is: n = 20, s^2 = 0.009

   $\sigma_0^2 = 0.010$

3. Test Statistic

$$\chi^2 = \frac{ns^2}{\sigma_0^2} = \frac{20 \times 0.009}{0.010} = 18$$

Degrees of freedom = n – 1 = (20 – 1)= 19

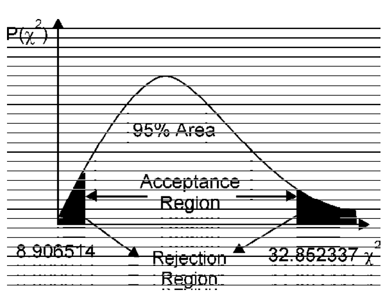


Table Value

At 5% LOS (selected), the table of c^2 for 19 D.F. in two tailed test is between (8.906514, 32.852377) (found from the tables in Appendix).

Conclusion

Since the computed value of c^2 = 18 is falling within the acceptance Region (limits obtained from the tables) for 19 D.F. at 5% LOS, we may accept H_0 and conclude that the sample evidence presents no reason to doubt the representative's claim.

7. A random sample of size 25 from a normal population gives a sample mean of 42 and a sample standard deviation of 6. Test the hypothesis that the population S.D. is 9. Clearly state the alternative hypothesis you allow for and the level of significance adopted.

Sol :

1. Null Hypothesis (H_0) : s^2 = 81

   Alternative Hypothesis (H_1) : s^2 > 81

2. The data given is: n = 25, s = 6, $\sigma_0^2 = (9)^2 = 81$

3. Test Statistic : $\chi^2 = \frac{ns^2}{\sigma_0^2} = \frac{25 \times 36}{81} = 11.11$

   Degrees of freedom = n – 1 = (25 – 1) = 24

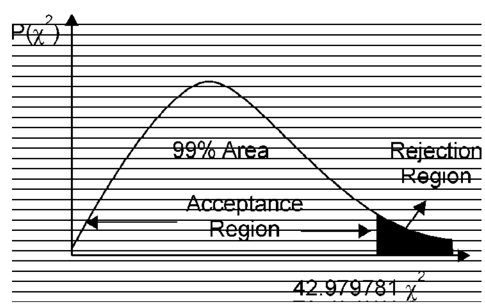


Table Value

At 1% LOS (selected), the table value of χ^2 for 24 D.F. (one tailed-right tail) is 42.979781

Conclusion

Since the computed value of c^2 = 11.11 is less than the table value of χ^2 = 42.979781 for 24 D.F. at 1% LOS, we may accept H_0 and conclude that the population S.D. is 9.

4.2.2 Test for Goodness of Fit

Q7. Explain briefly about test for goodness of fit?

Ans : **(Imp.)**

One of the very popular applications of χ^2 test is test of goodness of fit. It enables us to ascertain how the theoretical distributions such as Binomial, Poisson, Normal etc. can fit into empirical distributions obtained from sample data. When an ideal frequency curve whether normal or some other type is fitted to the data, we are interested out how well this curve fits with the observed facts.

A test of the concordance (goodness of fit) of the two can be made just by inspection, but such a test is obviously inadequate. Precision can be secured by applying the χ^2 test. The following are the steps in testing the goodness of fit :

1. Null and alternative hypotheses are established, and a significance level is selected for rejection of the null hypothesis.
2. A random sample of independent observations is drawn from a relevant statistical population.
3. A set of expected or theoretical frequencies is derived under the assumptions that the null hypothesis is true. This generally takes the form of assuming that a particular probability distribution is applicable to the statistical population under consideration.
4. The observed frequencies are compared with the expected, or theoretical frequencies.
5. If the calculated value of χ^2 is less than the table value at a certain level of significance (generally 5% level) and for certain degrees of freedom the fit is considered to be good, i.e., the divergence between the actual and expected frequencies is attributed to random fluctuations of sampling.

On the other hand, if the calculated value of χ^2 is greater than the table value, the fit is considered to be poor, i.e., it cannot be attributed to fluctuations of sampling rather it is due to the inadequacy of the theory to fit the observed facts.

Goodness of Fit

χ^2 test help us to find out how well the assumed theoretical distribution fit to the observed data. When some theoretical distribution is fitted to the given data, the statistician or managers will be interested in knowing as to how this distribution fits with the observed data.

This method of χ^2 test helps in answering this question.

If the calculated value of χ^2 is less than the table value at a certain level of significance, the fit is considered to be good one i.e., divergence between the observed and expected frequencies is attributable to fluctuations of sampling. But if the reverse occurs, the fit is not considered to be a good one. In short,

$$\chi^2_{cal} < \chi^2_{table} \Rightarrow \text{Good fit}$$

$$\chi^2_{cal} < \chi^2_{table} \Rightarrow \text{Not a good fit.}$$

If = 1, 2, ... n is a set of observed (experimental) frequencies and E_i (i = 1, ... n) is the corresponding set of theoretical frequencies then

$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$ with the condition that, $\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i = N$ = Total frequency follows, χ^2 - Distribution with (n – 1) d.o.f.

Steps for Test of Goodness of Fit

1. Null hypothesis : Good fit exists between the theoretical distribution and given data.
2. Alternative hypothesis : No good fit.
3. Level of significance is α.
4. Critical region : Reject null hypothesis if $\chi^2 > \chi^2_a$ with v d.o.f. i.e., theoretical distribution is a poor fit.
5. Computations : $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$
6. **Decision :** Accept null hypothesis, if $\chi^2 < \chi^2_{\alpha}$. i.e., the theoretical distribution is a good fit to the data.

Problems

8. Assume that Air Ticket reservations from Delhi to Gulf are uniformly distributed during all days in the winter season. To determine whether it is uniform we have selected a random sample of reservation lists for 10 days. The following information is drawn from the list.

Distribution of actual number of reservations.

Sl. No.	No. of Reservations
1	65
2	80
3	100
4	98
5	75
6	80
7	82
8	70
9	60
10	90

Test the validity of assumption using chi-square test.

Sol :

(i) Null Hypothesis

Air ticket reservations from Delhi to Gulf are uniformly distributed during all days in winter season.

Alternate Hypothesis

Air ticket reservations are not uniformly distributed during all days of winter season. A sample of 800 reservations for 10 days is given.

Therefore, expected reservation for each day is $\frac{800}{10} = 80$

(ii) Computing Test Statistic χ^2

Sl.No.	O	E	O – E	$(O-E)^2$	$(O-E)^2/E$
1	65	80	– 15	225	2.8125
2	80	80	0	0	0
3	100	80	20	400	5
4	98	80	18	324	4.05
5	75	80	-5	25	0.3125
6	80	80	0	0	0
7	82	80	2	4	0.05
8	70	80	– 10	100	1.25
9	60	80	– 20	400	5
10	90	80	10	100	1.25
	800			$\Sigma\left(\frac{(O-E)^2}{E}\right) = 19.725$	

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = 19.725$$

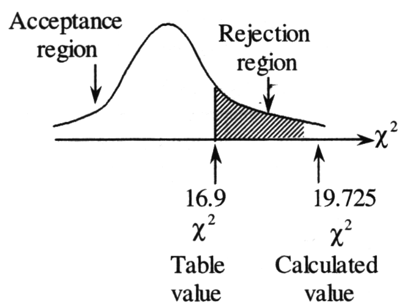
(iii) Level of significance, α = 0.0.5

Degrees of freedom, d.f = n – 1 = 10 – 1 = 9

(iv) Table value χ^2 for 9 d.f at 5% level of significance is 16.9

(v) Decision

The calculated value of χ^2 (19.725) is greater than table value 16.9. The null hypothesis is rejected. Hence, it can be said that air reservations from Delhi to Gulf are not uniformly distributed during all days in winter season.



Figure

9. A survey of 320 families with 5 children each revealed the following distribution :

Number of Boys	**5**	**4**	**3**	**2**	**1**	**0**
Number of Girls	**0**	**1**	**2**	**3**	**4**	**5**
Number of Families	**14**	**56**	**110**	**88**	**40**	**12**

Is this result consistent with the hypothesis is that male and female births are equally probable?

Sol :

Null Hypothesis (H_0) : Male and Female births are equally probable against

Alternative Hypothesis (H_1) : Male and Female births are not equally probable.

This assumptions of H_0 takes us to the probability of a male birth is p = $^1/_2$ and the underlying distribution is Binomial distribution. Since the birth can be a Male or Female the dichotomous classification.

Fix a = 5%.

Test Statistic : $\chi^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i}$ ~ a Chi Square Distribution with (k–1) D.F., where O_i refers to the observed frequencies and E_i refers to the expected frequencies. The probability of x male births in a family of 5 is given by

$P(x) = {}^5C_x\ P^x\ q^{5-x}$; for x = 0, 1, 2, 3, 4, 5,

$= {}^5C_x\ (1/2)^5$ [since p = q = 1/2]

To get the expected frequencies, multiply P(x) by the total number N = 320.

X	P(x)	Expected Frequency = N × P(x)
0	${}^5C_0\ (1/2)^5 = 1/32$	320 × 1/32 = 10
1	${}^5C_1\ (1/2)^5 = 5/32$	320 × 5/32 = 50
2	${}^5C_2\ (1/2)^5 = 10/32$	320 × 10/32 = 100
3	${}^5C_3\ (1/2)^5 = 10/32$	320 × 5/32 = 100
4	${}^5C_4\ (1/2)^5 = 5/32$	320 × 5/32 = 50
5	${}^5C_5\ (1/2)^5 = 1/32$	320 × 1/32 = 10

Arranging the observed and expected frequencies in the following table and calculating c^2.

O_i	E_i	$(O_i - E_i)^2$	$(O_i-E_i)^2/E_i$
14	10	16	1.60
56	50	36	0.72
110	100	100	1.00
88	100	144	1.44
40	50	100	2.00
12	10	4	0.40
			$\Sigma(O_i - E_i)^2 / E_i = 7.16$

Table Value

At 5% LOS (selected), the table value of c^2 for v = 6–1 = 5 at 5% LOS in two tailed test is between (0.831209), 12.832492) (found from the tables in Appendix).

Conclusion

Since the computed value of c^2 = 7.16 is falling within the acceptance region (limits obtained from the tables) for 5 D.F. at 5% LOS, we may accept H_0 and conclude that the male and female births are equally probable.

10. The following results are obtained when a dice is thrown 132 times :

Number Turned up	**1**	**2**	**3**	**4**	**5**	**6**
Frequency	**16**	**20**	**25**	**14**	**29**	**28**

Test the hypothesis is that the dice is unbiased.

Sol :

1) **Hypothesis (H_0):** The dice is unbiased

Against alternative Hypothesis (H_1): These dice is biased. This assumption of H_0 takes us to the probability of each face as $\left(\frac{1}{6}\right)$.

2) Fix ∞ = 2%

3) Test Statistic : $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ ~ a Chi-Square Distribution with (k–1) D.F., where O_i refers to the observed frequencies and E_i refers to the expected frequencies. The expected frequencies for each face is 132 × $\frac{1}{6} = 22$, with the assumption that the dice is unbiased.

Computation of the value of χ^2

Number turning up	Frequency Observed	Expected	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
O_i	E_i				
1	16	22	–6	36	36/22
2	20	22	–2	4	4/22
3	25	22	3	9	9/22
4	14	22	–8	64	64/22
5	29	22	7	49	49/22
6	28	22	6	36	36/22
				Total χ^2	198/22 = 9

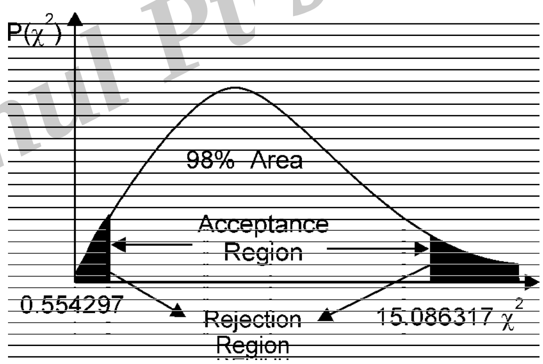


Table Value

At 2% LOS (selected), the table value of c² for n = 6 – 1 = 5 D.F. in two tailed test is between (0.554297, 15.086317) (found from the tables is Appendix).

Conclusion : Since the computed value of c² = 9 is falling within the acceptance Region (limits obtained from the tables) for 5 D.F. at 2% LOS, we may accept H_0 and conclude that the Dice is an unbiased one.

11. A die is rolled 100 times with the following distribution:

Number	1	2	3	4	5	6
Observed Frequency	**17**	**14**	**20**	**17**	**17**	**15**

At the 0.01 level of significance, determine whether die is true (or uniform).

Sol :

Chi-Square Test

Step-1

Null Hypothesis : (H_0) A die is true

Alternative Hypothesis (H_1): A die is not true.

Expected value (E) = $\frac{100}{6}$ = 16.67

Step-2

Computing test statistics,

$$\chi^2 = \Sigma \left[\frac{(O-E)^2}{E}\right]$$

Number	O	E	(O – E)	$(O - E)^2$	$\frac{(O-E)^2}{E}$
1	17	16.67	0.33	0.1089	0.0065
2	14	16.67	–2.67	7.1289	0.4276
3	20	16.67	3.33	11.0889	0.6652
4	17	16.67	0.33	0.1089	0.0065
5	17	16.67	0.33	0.1089	0.0065
6	15	16.67	–1.67	2.7889	0.1673
					$\Sigma\left[\frac{(O-E)^2}{E}\right]$ = 1.2796

$$\chi^2 = \Sigma \left[\frac{(O-E)^2}{E}\right] = 1.2796$$

Step-3

Level of Significance :

Level of significance = α = 0.01

Degree of freedom = n – 1

= 6 – 1

= 5

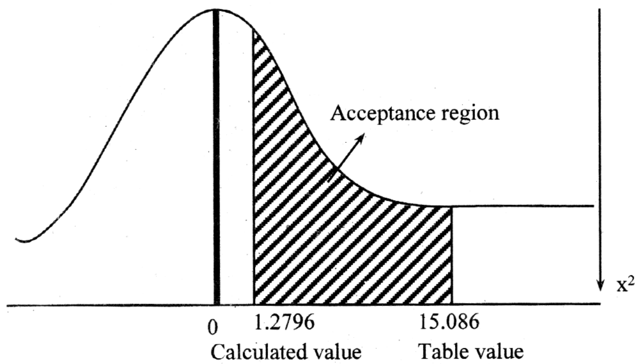
Step-4

Table Value:

Table value of χ^2 for 5 degree of freedom at 1% level of significance is 15.086

Step-5

Conclusion : As the calculated value of χ^2 is 1.2796 which is less than the table value i.e., 15.086. Hence, Null hypothesis is accepted, which states that the die is true (or uniform).



4.2.3 Test for Independence of Attributes

Q8. Explain test for independence of attribute.

Ans :

χ^2 test facilitates us to explain whether or not two attributes are associated. χ^2 is not a measure of the degree of relationship or the form of relationship between two attributes, but is simply a technique of judging the significance of such association or relationship between two attributes.

If the calculated value of χ^2 is less than table value, then null hypothesis is accepted which means the two attribute are independent and not associated. Otherwise, null hypothesis is not accepted which means the two attributes are associated and the association is not because of some chace factors, but it exists in reality.

In short,

$\chi^2_{cal} < \chi^2_{table} \Rightarrow$ Two attributes are independent

$\chi^2_{cal} > \chi^2_{table} \Rightarrow$ Two attributes are associated.

Analysis of R × C Table

The r × c tables are having r rows and c columns. There are two kinds of rxc table, in the first case the column totals (the sample sizes) are fixed, while in the second case only the grand total (the total for the entire table) is fixed.

As a result, there are also differences in the null hypothesis, we shall want to test.

In this problem r samples from one population with each item are classified with respect to two attributes. The row totals and column totals are not fixed, but random. Only the grand total N is fixed. The null hypothesis consist of testing whether the two attributes are independent. Then,

p_{ij} = (Problem of getting a value belonging to i^{ith} row)

× (Problem of getting a value belonging to j^{th} row)

The alternative hypothesis is that the two attributes are not independent.

The expected cell frequencies denoted by e are calculated by,

$$e_{ij} = \begin{cases} [(\text{Total observed frequencies in the } j^{th} \text{ column}) \times (\text{Total observed frequencies in the } r^{th} \text{ row})] \div \\ (\text{Total all cell frequencies}) \end{cases}$$

i.e., $e_{ij} = \dfrac{(\text{Column Total}) \times (\text{Row Total})}{\text{Grand Total}}$

Statistic for analysis of $r \times c$ table is,

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{C} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

With (r – 1) (c – 1) d.o.f.

Reject the null hypothesis if the calculated value of test statistic χ^2_{Cal} exceed χ^2_{α} with $(r-1)(C-1)$ d.o.f.

Reject Null Hypothesis if $\chi^2 > \chi^2_{\alpha}$

PROBLEMS

12. 1000 students at college level were graded according to their IQ and economic conditions of their home. Use Chi-square test to find out whether there is any association between economic condition at home and IQ.

Economic Condition	High IQ	Low IQ	Total
Rich	460	140	600
Poor	240	160	400
Total	700	300	1,000

Sol :

Null Hypothesis (H_0)

There is no association between economic conditions at home and IQ.

Alternate Hypothesis (H_1)

There is an association between economic conditions at home and IQ.

Computing Test Static

$$\chi^2 = \sum \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right]$$

O_{ij} = Observed frequency at the cell in the i^{th} row and j^{th} column

E_{ij} = Expected frequency at the cell in the ii^{th} row and j^{th} column.

Expected frequency $= \dfrac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$

i.e., $E_{ij} = \dfrac{RT \times CT}{GT}$ (Calculated for each box)

Calculation of Expected Frequencies

Economic Conditions	IQ		Total
	High	Low	
Rich	$\frac{600 \times 700}{1000} = 420$	$\frac{600 \times 300}{1000} = 180$	600
Poor	$\frac{400 \times 700}{1000} = 280$	$\frac{400 \times 300}{1000} = 120$	400
Total	700	300	G.T = 1000

Calculation of χ^2

O_{ij}	E_{ij}	$(O_{ij} - E_{ij})$	$(O_{ij} - E_{ij})^2$	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
460	420	40	1600	3.81
240	280	–40	1600	5.71
140	180	–40	1600	8.89
160	120	40	1600	13.33
				31.74

$$\chi^2 = \sum \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right]$$

$= 31.74$

Degree of freedom = (C – 1) (r – 1) = (2 – 1) (2 – 1) = (1 – 1) = 1

Level of significance = 5% (or) 0.05 [Assumed].

Table value of χ^2 is 1 d.f and 0.05 is 3.84

The calculated value of χ^2 is greater than the table value.

So, H_1 accepted, i.e., there is an association between economic conditions at home and IQ.

13. A researcher wants to know whether there exists any association between family expenditure an family size. A simple of 1000 families was taken and the following distribution is obtained.

Family size	Family expenditure		
	Low	Medium	High
3	200	100	100
4	150	120	30
5	100	80	120

Test the hypothesis at 1% level of significance.

Sol :

Null Hypothesis

H_o = The family expenditures are independent on family size.

Alternate Hypothesis

H_1 = The family expenditures are dependent on family size.

Computing Test Static

$$\chi^2 = \sum \left(\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right)$$

O_{ij} = Observed frequency of the cell in the i^{th} row and j^{th} column.

E_{ij} = Expected frequency of the cell in the i^{th} row and j^{th} column.

$$\text{Expected frequency} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

i.e., $E_{ij} = \dfrac{RT \times CT}{GT}$ (calculated for each box)

Calculation of Expected Frequencies

Family Size	Family Expenditure			
	Low	Medium	High	Total
3	$\frac{400 \times 450}{1000} = 180$ E_{11}	$\frac{400 \times 300}{1000} = 120$ E_{12}	$\frac{400 \times 250}{1000} = 100$ E_{13}	400
4	$\frac{300 \times 450}{1000} = 135$ E_{21}	$\frac{300 \times 300}{1000} = 90$ E_{22}	$\frac{300 \times 250}{1000} = 75$ E_{23}	300
5	$\frac{300 \times 450}{1000} = 135$ E_{31}	$\frac{300 \times 300}{1000} = 90$ E_{32}	$\frac{300 \times 250}{1000} = 75$ E_{33}	300
Total	**450**	**300**	**250**	**1000**

Calculation of χ^2

Group	O_{ij} (1)	E_{ij} (2)	$O_{ij} - E_{ij}$ (1) – (2) = (3)	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ $\frac{(3)^2}{2}$
11	200	180	20	2.22
12	100	120	–20	3.33
13	100	100	0	0
21	150	135	15	1.66
22	120	90	30	10
23	30	75	–45	27
31	100	135	–35	9
32	80	90	–10	1.11
33	120	75	45	27
			Total	**81.32**

$$\therefore \quad \chi^2 = \sum\left(\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right)$$

$$= 81.32$$

Level of significance, a = 0.01

Degree of freedom = (c – 1) (r – 1) = (3 – 1) (3 – 1)

= 2 × 2 = 4

Table value of χ^2 at 4 d.f and 0.01 is 13.277

Since calculated $\chi^2 - \chi^2_{table}$ (81.32 $\geq$ 13.277), H_1 is accepted at 1% level of significance i.e., Family expenditures are dependent on family size.

4.3 Correlation

Q9. Define Correlation. Explain the significance of Correlation.

Ans :

Meaning and Definition of Correlation

Correlation is the study of the linear relationship between two variables. When there is a relationship of 'quantitative measure between two set of variables, the appropriate statistical tool for measuring the relationship and expressing each in a precise way is known as correlation.

For example, there is a relationship between the heights and weights of persons, demand and prices of commodities etc.

Correlation analysis is the statistical tool we can use to describe the degree to which one variable is linearly related to another.

According to Croxton and Cowden, "The appropriate statistical tool for discovering and measuring the relationship of quantitative nature and expressing it in brief formula is known as correlation".

According to Tippet, "The effects of correlation are to reduce the range of uncertainty of our prediction".

The coefficient of correlation measures the degree of relationship between two set of figures. As the reliability of estimates depend upon the closeness of the relationship it is imperative that utmost care be taken while interpreting the value of coefficient of correlation, otherwise wrong conclusion can be drawn.

Significance of Measuring Correlation

1. Correction is very useful to economists to study the relationship between variables, like price and quantity demanded. To businessmen, it helps to estimate costs, sales, price and other related variables.
2. In economic theory we come across several types of variables which show some kind of relationship. For example, there exists a relationship between price, supply and quantity demanded; convenience, amenities, and service standards are related to customer retention; yield a crop related to quantity of fertilizer applied, type of soil, quality of seeds, rainfali and so on. Correlation analysis helps in measuring the degree of association and direction of such relationship.
3. The relation between variables can be verified and tested for significance, with the help of the correlation analysis. The effect of correlation is to reduce the range of uncertainty of our prediction.
4. The coefficient of correlation is a relative measure and we can compare the relationship between variables, which are expressed in different units.
5. Correlations are useful in the areas of healthcare such as determining the validity and reliability of clinical measures or in expressing how health problems are related to certain biological or environmental factors. For example, correlation coefficient can be used to determine the degree of inter-observer reliability for two doctors who are assessing a patient's disease.
6. Sampling error can also be calculated.
7. Correlation is the basis for the concept of regression and ratio of variation.
8. The decision making is heavily felicitated by reducing the range of uncertainty and hence empowering the predictions.

Q10. Explain various types of Correlation.

(or)

What are the different types of corelations.

Ans : **(Imp.)**

Broadly speaking, there are four types of correlation, namely,

A) Positive correlation,

B) Negative correlation,

C) Linear correlation and

D) Non-Linear Correlation.

A) Positive correlation

If the values of two variables deviate in the same direction i.e., if increase in the values of one variable results, on an average, in a corresponding increase in the values of the other variable or if a decrease in the values of one variable results, on an average, in a corresponding decrease in the values of the other variable, the corresponding correlation is said to be positive or direct.

Examples

i) Sales revenue of a product and expenditure on Advertising.

ii) Amount of rain fall and yield of a crop (up to a point)

iii) Price of a commodity and quantity of supply of a commodity

iv) Height of the Parent and the height of the Child.

v) Number of patients admitted into a Hospital and Revenue of the Hospital.

vi) Number of workers and output of a factory.

i) Perfect Positive Correlation : If the variables X and Y are perfectly positively related to each other then, we get a graph as shown in fig.below.

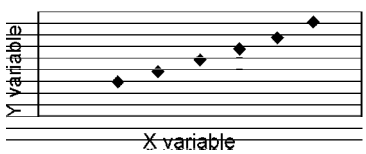


Fig.: Perfect Positive Correlation (r = +1)

ii) Very High Positive Correlation : If the variables X and Y are related to each other with a very high degree of positive relationship then we can notice a graph as in figure below.

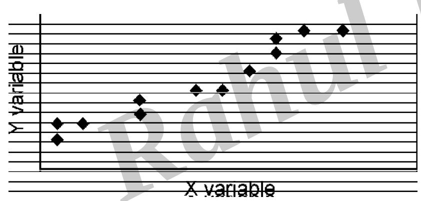


Fig.: Very High Positive Correlation (r = nearly + 1)

iii) Very Low Positive Correlation : If the variables X and Y are related to each other with a very low degree of positive relationship then we can notice a graph as in fig.below.

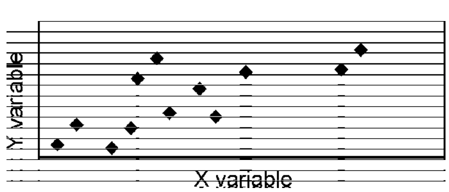


Fig.: Very Low Positive Correlation (r = near to +0)

B) Negative Correlation

Correlation is said to be negative or inverse if the variables deviate in the opposite direction i.e., if the increase (decrease) in the values of one variable results, on the average, in a corresponding decrease (increase) in the values of the other variable.

Examples

1. Price and demand of a commodity.
2. Sales of Woolen garments and the day temperature.

i) Perfect Negative Correlation : If the variables X and Y are perfectly negatively related to each other then, we get a graph as shown in fig.below.

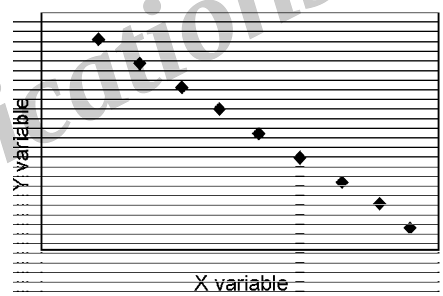


Fig.: Perfect Negative Correlation (r = –1)

ii) Very High Negative Correlation : If the variables X and Y are related to each other with a very high degree of negative relationship then we can notice a graph as in fig.below.

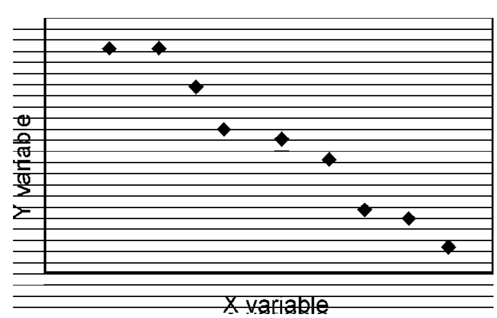


Fig.: Very High Negative Correlation (r = near to –1)

iii) Very low Negative Correlation : If the variables X and Y are related to each other with a very low degree of negative relationship then we can notice a graph as in fig.below.

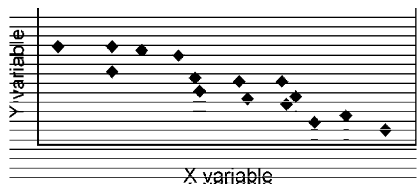


Fig.: Very Low Negative Correlation

(r = near to 0 but negative)

iv) No Correlation : If the scatter diagram show the points which are highly spread over and show no trend or patterns we can say that there is no correlation between the variables.

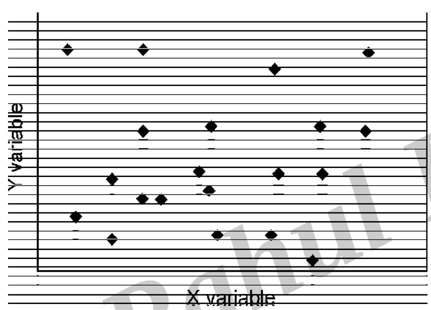


Fig.: No Correlation (r = 0)

C) Linear Correlation

Two variables are said to be linearly related if corresponding to a unit change in one variable there is a constant change in the other variable over the entire range of the values.

If two variables are related linearly, then we can express the relationship as

$$Y = a + b X$$

where '**a**' is called as the "intercept" (If X = 0, then Y = a) and 'b' is called as the "rate of change" or slope.

If we plot the values of X and the corresponding values of Y on a graph, then the graph would be a straight lines as shown in fig.below.

Example

X	1	2	3	4	5
Y	8	11	14	17	20

For a unit change in the value of x, a constant 3 units changes in the value of y can be noticed. The above can be expressed as : Y = 5 + 3x.

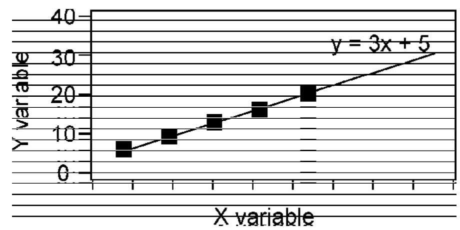


Fig.: Linear Correlation

D) Non Linear (Curvilinear) Correlation

If corresponding to a unit change in one variable, the other variable does not change in a constant rate, but change at varying rates, then the relationship between two variables is said to be nonlinear or curvilinear as shown in fig.below. In this case, if the data are plotted on the graph, we do not get a straight line curve.

Mathematically, the correlation is non-linear if the slope of the plotted curve is not constant. Data relating to Economics, Social Science and Business Management do exhibit often non-linear relationship. We confine ourselves to linear correlation only.

Example

X	–3	–2	–1	0	1	2	3
Y	9	4	1	0	1	4	9

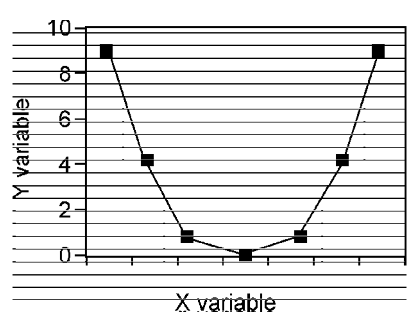


Fig.: Non Linear Correlation

Q11. Explain the Properties of Correlation.

Ans :

i) The value of correlation 'r' varies between [–1, +1]. This indicates that the r values does not exceed unity.

ii) Sign of the correlation sign of the Covariance.

iii) If r = –1 variables and perfectly negatively correlated.

iv) If r = + 1 variables are perfectly positively correlated.

If r = 0 variables are not correlated in a linear fashion. There may be non-linear relationship between variables.

Correlation coefficient is independent of change of scale and shifting of origin. In other words, Shifting the origin and change the scale do not have any effect on the value of correlation.

4.3.1 Limits for Coefficient of Correlation

Q12. State the Limits for Coefficient of Correlation.

Ans :

Limits for Co-efficient of Correlation for (x, y)

The value of the coefficient of correlation should lie between +1 and –1. If r = +1, the correlation is perfect and positive and if r = –1 the correlation is perfect and negative. When r = 0, it means that there is no relationship between the two variables.

Hence, $-1 < r(x, y) \leq 1$

Note

The correlation coefficient describes not only the magnitude of correlation but also its direction. Thus, +1 would mean that correlation is positive and the magnitude of correlation is 1.

Similarly -1 means correlation is negative and the magnitude of correlation is again 1.

	Degree of Correlation	The Value of r(x,y) (Positive and Negative)
1.	Perfect correlation	Exactly 1
2.	Very high degree of correlation	0.90 and above but less than 1
3.	Fairly high degree of correlation	0.75 and abovebut less than 0.90
4.	Moderate degree of correlation	0.50 and above but less than 0.75
5.	Low degree of correlation	0.25 and above but less than 0.50
6.	Very high degree of correlation	Below 0.25
7.	Absence of correlation	Equal to 0

Q13. Explain various methods of Correlation.

Ans :

The various method of studying linear correlation can be shown diagrammatically as follows:

1. Scatter Diagram
2. Two way frequency table

3. Spearman's Rank Correlation Method
4. Method of Concurrent Deviations
5. Method of Least Squares.

Q14. Define scatter diagram. Explain merits and demerits of scatter diagram.

Ans : **(Nov.-20, Imp.)**

Scatter diagram is a special type of dot chart that is used for graphical representation of the relationship between two quantitative variables.

Merits of Scatter Diagram

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

Demerits of Scatter Diagram

1. The major limitation of the method is that it only gives a visual picture of the relationship of two variables. It only tells us whether there is correlation between the variables, and if so, then in which direction, positive or negative.
2. It does not give an idea about the precise degree of relationship as it is not amenable to mathematical treatment.

Q15. What is Two way frequency table ?

Ans :

This is the simplest method of judging association between two variables. The two-way frequency table is prepared by indicating one variable in the rows (horizontal) and the other variable in the columns (vertical).

The frequencies are shown in the respective squares which are equal to m × n where m and n are the respective sizes of classes of rows and columns. The nature of concentration of frequencies in various square reveals the type of correlation, the table gives only a very rough idea and it is difficult to quantitatively indicate the degree of correlation.

Therefore, it is not considered to be a scientific method. Given below is a two-way frequency table dealing with marks in Telugu in the rows and marks in English in the columns.

Marks in		**Marks in English**			**Total**	
Telugu	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**	
0-10	5	8				13
10-20		6	16			22
20-30				6	4	10
30-40				3	3	6
40-50				3	4	7
Total	**5**	**14**	**16**	**12**	**11**	**58**

The scatter of values on top left to bottom right reveals that lower marks in economic are associated with lower marks in statistics.

Q16. What is Concurrent Deviations Method?

Ans :

In this method, only the direction of change c_x, c_y in the concerned variables is taken into account. For each term the change is considered with reference to the previous value, which may be either plus (+) or minus (–). The formula used for concurrent deviation method is,

$$r = \pm \sqrt{\pm\left[\frac{(2c - n)}{n}\right]}$$

Where,

r = Coefficient of concurrent deviations

c = Number of concurrent deviations (i.e., the number of positive signs in (c_x, c_y)

n = Number of pairs of deviations compared.

4.4 Karl Pearson's Coefficient of Correlation

Q17. What is Karl Pearson's Coefficient of Correlation? Explain properties of Coefficient of Correlation.

Ans : **(Imp.)**

Karl Parson's Coefficient of Correlation is arrived at with the help of a statistical formula that takes into account the mean and standard deviation of the two variables, the number of such observations and the covariance between them. Since Karl Pearson's coefficient of correlation is a number, it can describe the strength of the correlation in greater detail and more objectively. A value of –1 signifies "absolute" negative correlation, a value between –1 and –0.5 signifies strong negative correlation, a value between –0.5 and –0.25 signifies moderate negative correlation and a value between –0.25 and 0 signifies weak negative correlation. Similarly, a value of +1 signifies "absolute" positive correlation, a value between +1 and +0.5 signifies strong positive correlation, a value between +0.5 and +0.25 signifies moderate positive correlation and a value between +0.25 and 0 signifies weak positive correlation

Properties of Karl Pearson's Coefficient of Correlation

1. It is based on Arithmetic Mean and Standard Deviation.
2. It lies between –1
3. It measures both direction as well as degree of change. If r is less than 0, there is negative correlation, which means the direction of change of the two variables will be opposite. If r is more than 0, there is positive correlation, which means the direction of change of the two, variables will be same. Higher the value of r, greater is the degree of correlation. Hence, Karl Pearson's coefficient of correlation is said to be the ideal measure of correlation.
4. It is independent of change in scale. In other words, if a constant amount is added/ subtracted from all values of a variable, the value of r does not change.
5. It is independent of change in origin. Thus, if a constant amount is multiplied with or divides all values of a variable, *Y* does not change.
6. It is independent of direction. In other words, Correlation of X and Y is same as Correlation of Y and X.
7. It is a pure number without any units. In other words, it is independent of the unit of measurement of the 2 variables.
8. It takes into account all items of the variable(s).
9. It does not prove causation but is simply a measure of co-variation.
10. Correlation coefficient of two variables X and Y is the Geometric Mean of two regression coefficients, regression coefficient of X on Y and regression coefficient of Y on X. Symbolically,

    r = Square root of (b_{xy} * b_{yx})

11. Correlation coefficient can be calculated between two unrelated variables and such a number can be misleading. Such correlation is called accidental correlation, spurious correlation or non sense correlation.

Q18. Explain merits and demerits of Coefficient of Correlation.

Ans :

Merits of Karl Pearson's Coefficient of Correlation

1. It takes into account all items of the variable(s).
2. It is a numerical measure and hence more objective.
3. It measures both direction as well as degree of change.
4. It facilitates comparisons between two series.
5. It is capable of further Algebraic treatment
6. It is more practical and hence popular and is more commonly used.

Demerits of Karl Pearson's Coefficient of Correlation

1. It is not easy to calculate as complex formulae are involved.
2. It is more time consuming compared to methods such as rank correlation
3. It assumes a linear relationship between the two variables which may not be correct
4. It is impacted by extreme values as it is based o mean and standard deviation.
5. It is not easy to interpret.

Q19. Explain the methods of Coefficient of Correlation.

Ans :

i) Direct Method when deviations are taken from actual mean

$$r = \frac{\Sigma xy}{N\sigma_x \alpha_y}$$

However, this formula is transformed in the following form

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

Where

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

Steps :

1. Find the means of the two series ($\bar{X}$, $\bar{Y}$)
2. Take the deviations of X series from the mean of X and denote these deviations as x.
3. Square these deviations and obtain the total. Denote it as Σx^2.
4. Take the deviations of Y series from the Mean of Y and denote these deviations as y.

5. Square these deviations, obtain the total and denote it as Σy^2.

6. Multiply the deviations of X and Y series, obtain the total and denote it Σxy.

7. Substitute the above values in the formula.

ii) Short-Cut Mfethod

When deviations are taken from assumed mean.

When actual mean is in fraction, then the above formula becomes tedious. In such cases, assumed mean is used for calculating correlation. The formula is.

$$r = \frac{\Sigma dxdy - \frac{\Sigma dx.\Sigma dy}{N}}{\sqrt{\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}}\sqrt{\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}}}$$

Where

$\Sigma dxdy$ = Sum of the product of the deviations of X and Y series from their assumed means.

Σdx^2 = Sum of the squares of the deviations of X series from an assumed mean.

Σdy^2 = Sum of the squares of the deviations of Y series from an assumed mean.

Σdx = Sum of the deviations of X series from an assumed mean.

Σdy = Sum of the deviations of Y Series from an assumed mean.

N = No. of Pairs of observations.

The values of coefficient of correlation as obtained by above formulae will always lie between ± 1. When there is perfect positive correlation its value is $+1$ and when there is perfect negative correlation, its value is –1. When $r = 0$ means that there is no relationship between the two variables. We normally get values which lie between $+1$ and –1.

Probable Error of the Coefficient of correlation and its interpretation.

The probable error of the coefficient of correlation helps in interpretation. The probable error of the coefficient of correlation is obtained as follows:

$$\text{P.E. of r} = 0.6745\ \frac{1 - r^2}{\sqrt{N}}$$

Where

r = Coefficient of correlation;

N = Number of pairs of observations.

If the probable error is added to and subtracted from the coefficient of correlation, it would give two such limits within which we can reasonably expect the value of coefficient of correlation to vary.

Symbolically P (rho) = r $\pm$ P.E.

Where 'P' denotes the correlation in the population. Suppose, the Coefficient of correlation for a pair of 10 observations is 0.8 and its P.E. is 0.05. the limits of the correlation in the population would be r $\pm$ P.E. i.e. 0.8 $\pm$ 0-.05 or 0.75 – 0.85.

If the value of r is less than the probable error then r is not at all significant, i.e. there is no evidence of correlation. If the value of r is more than six times the probable error, it is significant. Hence it can be said that r is significant, when

$$r > 6 \text{ P.E. or } \frac{r}{\text{P.E}} > 6$$

Problems

14. Find the coefficient of correlation between x and y

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

Sol :

X	Y	$x = X - \bar{X}$	x^2	$y = Y - \bar{Y}$	y^2	xy
65	67	–3	9	–2	4	6
66	68	–2	4	–1	1	2
67	65	–1	1	–4	16	4
67	68	–1	1	–1	1	1
68	72	0	0	3	9	0
69	72	1	1	3	9	3
70	69	2	4	0	0	0
72	71	4	16	2	4	8
544	552	$\Sigma x = 0$	$\Sigma x^2 = 36$	$\Sigma y = 0$	$\Sigma y^2 = 44$	$\Sigma xy = 24$

$$\bar{X} = \frac{\Sigma x}{n} = \frac{544}{8} = 68, \qquad \bar{Y} = \frac{\Sigma y}{n} = \frac{552}{8} = 69$$

As per Karl Perason, coefficient of correlation

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{24}{\sqrt{36 \times 44}} = +0.603$$

15. Calculate Karl Pearson's Coefficient of correlation between expenditure on advertising and sales from the data given below

Advertising Expenses ('000 Rs.)	39	65	62	90	82	75	25	98	36	78
Sales (Lakh Rs.)	47	53	58	86	62	68	60	91	51	84

Sol :

Calculation of Coefficient of Correlation

Advertising X	X- $\bar{X}$ x	x^2	Sales Y	Y- $\bar{Y}$ y	y^2	x y
39	-26	676	47	-19	361	+494
65	0	0	53	-13	169	0
62	-3	9	58	-8	64	+24
90	+25	625	86	+20	400	+500
82	+17	289	62	-4	16	-68
75	+10	100	68	+2	4	+20
25	-40	1600	60	-6	36	+240
98	+33	1089	91	+25	625	+825
36	-29	841	51	-15	225	+435
78	+13	169	84	+18	324	+234
ΣX=650	**Σx=0**	**Σx²=5398**	**Σy=660**	**Σy=0**	**Σy²=2224**	**Σxy=2704**

Advertising Cost $\bar{X} = \dfrac{\Sigma X}{N} = \dfrac{650}{10} = 65;$

Sales $\bar{Y} = \dfrac{\Sigma X}{X} = \dfrac{660}{10} = 66$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}};$$

$\Sigma xy = 2704, \Sigma x^2 = 5398, \Sigma y^2 = 2224$

$$= \frac{2704}{\sqrt{5398 \times 2224}}$$

$$= \frac{2704}{1200 \quad 5152} = +0.7804$$

16. Calculate the coefficient of correlation for the following age of husbands and wives.

Husband's Age (in years)	**23**	**27**	**28**	**29**	**30**	**31**	**33**	**35**	**36**	**39**
Wife's Age (in years)	**18**	**22**	**23**	**24**	**25**	**26**	**28**	**29**	**30**	**32**

Sol :

Age of Husband	X–A	Age of Y–A				
X	A=30 dx	Wife dx²	Y	A=25 dy	dy²	dxdy
23	–7	49	18	–7	49	+49
27	–3	9	22	–3	9	+9
28	–2	4	23	–2	4	+4
29	–1	1	24	–1	1	+1
30	0	0	25	0	0	+0
31	+1	1	26	+1	1	+1
33	+3	9	28	+3	9	+9
35	+5	25	29	+4	16	+20
36	+6	36	30	+5	25	+30
39	+9	81	32	+7	49	+63
	Σdx = +11	**Σdx² = 215**		**Σdy = +7**	**Σdy² = 163**	**Σdxdy = 186**

$$r = \frac{\Sigma dxdy - \dfrac{\Sigma dx.\Sigma dy}{N}}{\sqrt{\Sigma dx^2 - \dfrac{(\Sigma dx)^2}{N}}\sqrt{\Sigma dy^2 - \dfrac{(\Sigma dy)^2}{N}}}$$

$\Sigma dxdy = 186$, $\Sigma dx = 11$, $\Sigma dy = +7$, $\Sigma dx^2 = 215$, $\Sigma dy^2 = 163$, $N = 10$

$$\frac{185 - \dfrac{11 \times 7}{10}}{\sqrt{215 - \dfrac{(11)^2}{10}}\sqrt{163 - \dfrac{(7)^2}{10}}}$$

$$\frac{186 - 7.7}{\sqrt{215 - 12.1}\sqrt{163 - 4.9}} \quad \frac{178.3}{14.244 \times 12.574} = \frac{178.3}{179.104} = 0.966$$

There is a very high degree of positive correlation between the ages of husbands and wives.

17. Given

Total of Product of deviations of X and Y series = +85

Number of pairs of observations = 10

Total of the deviations of X series = +6

Total of the deviations of Y series = +13

Total of the square of deviations of X series 162

Total of the square of deviations of Y series 211

Find out the coefficient of correlation when arbitrary means of X series and Y series are 35 and 30 respectively.

Sol :

Given

$\Sigma dxdy = 85,\ \Sigma dx = 6,\ \Sigma dy = 13,\ \Sigma dx^2 = 162,\ \Sigma dy^2 = 211,\ N = 10$

Karl Pearson's Coefficient of Correlation

$$r = \frac{\Sigma dx.dy - \frac{\Sigma dx.dy}{N}}{\sqrt{(\Sigma dx^2 \frac{(\Sigma dx^2)}{N}}\sqrt{\Sigma dy^2 \frac{(\Sigma dy)^2}{N}}} = \frac{85 - \frac{6 \times 13}{10}}{\sqrt{162 - \frac{(6)^2}{10}}\sqrt{211 - \frac{(13)^2}{10}}}$$

$$= \frac{85 - 7.8}{\sqrt{162 - 3.6}\sqrt{211 - 16.9}} = \frac{77.2}{\sqrt{158.4} \times \sqrt{194.1}} = \frac{77.2}{12.586 \times 13.932} = \frac{772}{175.348}$$

$= +0.44$

18. Find Karl Pearson's coefficient of correlation from the following data and comment on the value of the coefficient.

Cost	39	65	62	90	82	75	25	98	36	78
Sales	47	53	58	86	62	68	60	91	51	84

Sol :

Let X be the cost and Y be sales averages of X and Y

$$\bar{X} = \frac{\Sigma X}{N} = \frac{650}{10} = 65$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{660}{10} = 66$$

Karl Pearson's coefficient of correlation,

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$$

Where,

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

'r' is independent of change in origin and scale.

Origin change: Deviations from mean; Scale change dividing by any number.

X	Y	(X– 65) (x)	(Y – 66) (y)	x^2	y^2	xy
39	47	–26	–19	676	361	494
65	53	0	–13	0	169	0
62	58	–3	–8	9	64	24
90	86	25	20	625	400	500
82	62	17	–4	289	16	-68
75	68	10	2	100	4	20
25	60	–40	–6	1600	36	240
98	91	33	25	1089	625	825
36	51	–29	–15	841	225	435
78	84	13	18	169_	324	234
ΣX=650	**ΣY=660**			**x^2=5398**	**y^2=2224**	**xy=2704**

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} = \frac{2704}{\sqrt{5398}\sqrt{2224}} = \frac{2704}{73.47 \times 47.16} = \frac{2704}{3464.84} = 0.780$$

Karl Pearson's coefficient correlation = 0.780

Conclusion

There is a positive correlation between the variables X and Y.

19. Calculate Karl Pearson's Coefficient of Correlation for the following data :

X	78	89	96	69	59	79	68	62
Y	121	137	156	112	107	136	123	108

Ans :

X	X–A=dx	dx^2	Y	Y–A=dy	dy^2	dxdy
78	9	81	121	9	81	81
89	20	400	137	25	625	500
96	27	729	156	44	1936	1188
69 A	0	0	112 A	0	0	0
59	–10	100	107	–5	25	50
79	10	100	136	24	576	240
68	–1	1	123	11	121	–11
62	–7	49	108	–4	16	28
	48	1460		104	3380	2076

$$r = \frac{\sum dxdy - \frac{\sum dx.dy}{N}}{\sqrt{\sum dx^2 - \left(\frac{\sum dx}{N}\right)^2}\sqrt{\sum dy^2 - \left(\frac{\sum dy}{N}\right)^2}}$$

$$r = \frac{2076 - \frac{48 \times 104}{8}}{\sqrt{1460 - \frac{(48)^2}{8}}\sqrt{3380 - \frac{(104)^2}{8}}} = \frac{2076 - \frac{4992}{8}}{\sqrt{1460 - \frac{2304}{8}}\sqrt{3380 - \frac{10816}{8}}}$$

$$r = \frac{2076 - 624}{\sqrt{1460 - 288}\sqrt{3380 - 1352}} = \frac{1452}{\sqrt{1172}\sqrt{2028}}$$

$$r = \frac{1452}{34.234 \times 45.033} = \frac{1452}{1541.66}$$

$$r = 0.9418$$

There is a high positive correlation between y and x.

20. Find the correlation coefficient between age and playing habits of the following students:

Age	**15**	**16**	**17**	**18**	**19**	**20**
Playing habits	**80**	**75**	**60**	**40**	**30**	**15**

Ans :

X	$\mathbf{X - \bar{X} = x}$	**Y**	$\mathbf{Y - \bar{Y} = y}$	**xy**	$\mathbf{x^2}$	$\mathbf{y^2}$
15	–2.5	80	30	–75	6.25	900
16	–1.5	75	25	37.5	2.25	625
17	–0.5	60	10	–5	0.25	100
18	0.5	40	–10	–5	0.25	100
19	1.5	30	–20	–30	2.25	400
20	2.5	15	–35	–87.5	6.25	1225
105	0	300	0	240	17.50	3350

i) $\bar{X} = \frac{\sum X}{N} = \frac{105}{6} = 17.5$

$\bar{Y} = \frac{\sum Y}{N} = \frac{300}{6} = 50$

ii) $$r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}} = \frac{240}{\sqrt{17.50 \times 3350}}$$

$$r = \frac{240}{\sqrt{58,625}} = \frac{240}{242.126} = 0.9912.$$

4.4.1 Spearman's Rank Correlation

Q20. Explain about rank correlation and its features.

Ans :

The Karl Pearson's method is based on the assumption that the population being studied is normally distributed. When it is known that the population is not normal or when the shape of the distribution is not known, there is need for a measure of correlation that involves no assumption about the parameter of the population.

It is possible to avoid making any assumptions about the populations being studied by ranking the observations according to size and basing the calculations on the ranks rather than upon the original observations. It does not matter which way the items are ranked, item number one may be the largest or it may be the smallest. Using ranks rather than actual observations gives the coefficient of rank correlation.

This method of finding out covariability or the lack of it between two variables was developed by the British Psychologist Charles Edward Spearman in 1904. This measure is especially useful when quantitative measures for certain factors (such as in the evaluation of leadership ability or the judgment of female beauty) cannot be fixed, but the individual in the group can be arranged in order thereby obtaining for each individual a number indicating his (her) rank in the group. Spearman's rank correlation coefficient is defined as :

$$R = 1 - \frac{6\Sigma D^2}{N(N^2-1)} \text{ or } 1 - \frac{6\Sigma D^2}{N^3 - N}$$

where R denotes rank coefficient of correlation and D refers to the difference of rank between paired items in two series.

Features of Spearman's Correlation Coefficient

- The sum of the differences of ranks between two variables shall be zero. Symbolically. Xd = 0
- Spearman's correlation coefficient is distribution-free or non-parametric because no strict assumptions are made about the form of population from which sample observations are drawn.
- The Spearman's correlation coefficient is nothing but Karl Pearson's correlation coefficient between the ranks. Hence, it can be interpreted in the same manner as Pearsonian correlation coefficient.

  In rank correlation we may have two types of problems :

  - Where ranks are given.
  - Where ranks are not given.

i) Where Ranks are Given

Steps

(i) Take the differences of the two ranks, i.e. $(R_1 - R_2)$ and denote these differences by D.

(ii) Square these differences and obtain the total ΣD^2.

(iii) Apply the formula $R = 1 - \frac{6\Sigma D^2}{N^3 - N}$

Example

The ranking of 10 students in two subjects A and B are as follows :

A	B	A	B
6	3	4	6
5	8	9	10
3	4	7	7
10	9	8	5
2	1	1	2

Calculate rank correlation coefficient.

Sol. :

Calculation of Rank Correlation Coefficient

R_1	R_2	D	$(R_1 - R_2)^2$ D^2
6	3	3	9
5	8	–3	9
3	4	–1	1
10	9	1	1
2	1	1	1
4	6	– 2	4
9	10	– 1	1
7	7	–0	0
8	5	3	9
1	2	–1	1
			$\Sigma D^2 = 36$

$$R = 1 - \frac{6\Sigma D^2}{N^3 - N} = 1 - \frac{6 \times 36}{10^3 - 10}$$

$$= 1 - \frac{216}{990} = 0.782$$

Where Ranks are Not Given

Steps :

1) to assign the ranks.

Note : Ranks can be as-signed by taking either highest value as 1 or the lowest value as 1. But whether we start with the lowest value or the highest value we must follow the same method in case of both the variables.

2) Take the differences of the two ranks, i.e. $(R_1 - R_2)$ and denote these differences by D.

3) Square these differences and obtain the total ΣD^2.

4) Apply the formula $R = 1 - \frac{6\Sigma D^2}{N^3 - N}$

Example

Quotations of Index Numbers of security prices of a certain joint stock company are given below :

Year	Debenture price	Share price
1	97.8	73.2
2	99.2	85.8
3	98.8	78.9
4	98.3	75.8
5	98.4	77.2
6	96.7	87.2
7	97.1	83.8

Using rank correlation method, determine the relationship between debenture prices and share prices.

Sol :

Calculation of Rank Correlation Coefficient

X	Rx	Y	Ry	$(R_x - R_y)$ D	$(Rx - Ry)^2$ D^2
97.8	3	73.2	1	2	4
99.2	7	85.8	6	1	1
98.8	6	78.9	4	2	4
98.3	4	75.8	2	2	4
98.4	5	77.2	3	2	4
96.7	1	87.2	7	–6	36
97.1	2	83.8	5	–3	9
					$\Sigma D^2 = 62$

$$R = 1 - \frac{6\,\Sigma\,D^2}{N^3 - N} = 1 - \frac{6 \times 62}{7^3 - 7} = 1 - 1.107 = -0.107.$$

Equal Ranks

In some cases it may be found necessary to rank two or more individuals or entries as equal. In such a case it is customary to give each individual an average rank. Thus, if two individuals are ranked equal at fifth place, they are each given the rank $\frac{5+6}{2}$. that is 5.5 while, if three are ranked equal at fifth place, they are given the rank $\frac{5+6+7}{3} = 6$. In other words, where two or more items are to be ranked equal, the rank assigned for purposes of calculating coefficient of correlation is the average of the ranks which these individuals would have got had they differed slightly from each other.

Where equal ranks are assigned to some entries an adjustment in the above formula for calculating the rank coefficient of correlation is made.

The adjustment consists of adding $\frac{1}{12}(m^3 - m)$ to the value of ΣD^2, where M stands for the number of items whose ranks are common. If there are more than one such group of items with common rank, this value is added as many times the number of such groups. The formula can thus be written

$$R = 1 - \frac{6\left\{\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + ...\right\}}{N^3 - N}$$

Example

Obtain the rank correlation coefficient between the variables X and Y from the following pairs of observed values.

X:	50	55	65	50	55	60	50	65	70	75
Y:	110	110	115	125	140	115	130	120	115	160

Sol :

For finding ranks correlation coefficient first rank two various values. Taking lowest as 1 and next higher as 2, etc..

X	Rank X R_1	Y	Rank YR_2 R_2	D	$(R_1 - R_2)^2$ D^2
50	2	110	1.5	–0.5	0.25
55	4.5	110	1.5	3.0	9.00
65	7.5	115	4	3.5	12.25
50	2	125	7	–5	25.00
55	4.5	140	9	–4.5	20.25
60	6	115	4	2	4.00
50	2	130	8	6	36.00
65	7.5	120	6	1.5	2.25
70	9	115	4	5.00	25.00
75	10	160	10	0.00	00.00
					$\Sigma D^2 = 134$

It may be noted that in series X, 50 has repeated thrice (m = 3), 55 has been repeated twice (m = 2), 65 has been repeated twice (m = 2). In series V, 110 has been repeated twice (m = 2) and 115 thrice (m = 3).

$$R = 1 - \frac{6\left\{\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right\}}{N^3 - N}$$

$$R = 1 - \frac{6\left\{134 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^3 - 3)\right\}}{10^3 - 10}$$

$$= 1 - \frac{6[134 + 2 + .5 + .5 + .5 + 2]}{990}$$

$$= 1 - \frac{6(139.5)}{990} = 1 - \frac{837}{990} = 1 - .845 = 0.155$$

Q21. What are the merits and demerits of rank correlation ?

Ans :

Merits

1. It is easy to calculate and understand as compared to Pearson's *r.*
2. When the ranks of different values of the variables are given, it is then the only method left to calculate the degree of correlation.
3. When actual values are given and we are interested in using this formula then, we have to give ranks to calculate correlation.
4. This method is employed usefully when the data is given in a qualitative nature like beauty, honesty, intelligence etc.

Demerits

1. This method cannot be employed in a grouped frequency distribution.
2. If the items exceed 30, it is then difficult to find out ranks and their differences.
3. This method lacks precision as compared to Pearson's co- efficient of correlation, as all the information concerning the variables is not used. It is just possible that the difference between r_k and r may be very insignificant.

PROBLEMS ON RANK CORRELATION

21. Twelve entries in painting competition were ranked by two judges as shown below :

Judge	**1**	**5**	**2**	**3**	**4**	**1**	**6**	**8**	**7**	**10**	**9**	**12**	**11**
Judge	**2**	**4**	**5**	**2**	**1**	**6**	**7**	**10**	**9**	**11**	**12**	**3**	**8**

Ans :

Judge 1	Judge 2	$R_1 - R_2 = d$	d^2
5	4	1	1
2	5	–3	9
3	2	1	1
4	1	3	9
1	6	–5	25
6	7	–1	1
8	10	–2	4
7	9	–2	4
10	11	–1	1
9	12	–3	9
12	3	9	81
11	8	3	9
			154

$$r_s = 1 - \frac{6\sum d^2}{(N)^3 - N}$$

$$r_s = 1 - \frac{6 \times 154}{(12)^3 - 12}$$

$$r_s = 1 - \frac{924}{1728 - 12}$$

$$r_s = 1 - \frac{924}{1716}$$

$$r_s = 1 - 0.5385$$

$$r_s = 1 - 0.4615 \text{ or } 0.46$$

22. Ten competitors in a beauty contest are ranked by three judges in the following order.

Ist Judge	**1**	**5**	**4**	**8**	**9**	**6**	**10**	**7**	**3**	**2**
2nd Judge	**4**	**8**	**7**	**6**	**5**	**9**	**10**	**3**	**2**	**1**
3rd Judge	**6**	**7**	**8**	**1**	**5**	**10**	**9**	**2**	**3**	**4**

Use the rank correlation coefficient to discuss which pair of judges has the nearest approach to beauty.

Ans :

Judge 1 (R_1)	Judge 2 (R_2)	Judge 3 (R_3)	R_1–R_2 D_1	R_2–R_3			R_1–R_3	
				D_1^2	D_2	D_2^2	D_3	D_3^2
1	4	6	–3	9	–2	4	–5	25
5	8	7	–3	9	1	1	-2	4
4	7	8	–3	9	1	1	–4	16
8	6	1	2	4	5	25	7	49
9	5	5	4	16	0	0	4	16
6	9	10	–3	9	–1	1	–4	16
10	10	9	0	0	1	1	1	1
7	3	2	4	16	1	1	5	25
3	2	3	1	1	–1	1	0	0
2	1	4	1	1	–3	9	–2	4
				74		44		156

i) Between 1st and 2nd judge

$$r_s = 1 - \frac{6\sum(d)^2}{N^3 - N} = 1 - \frac{6(74)}{(10)^3 - 10}$$

$$r_s = 1 - \frac{444}{1000 - 10}$$

$$r_s = 1 - \frac{444}{990}$$

$$r_s = 1 - 0.4485$$

$$r_s = 0.5515$$

ii) Between 2nd and 3rd judge

$$r_s = 1 - \frac{6\sum(d)^2}{N^3 - N}$$

$$r_s = 1 - \frac{6 \times (44)}{(10)^3 - 10}$$

$$r_s = 1 - \frac{264}{1000 - 10}$$

$$r_s = 1 - \frac{264}{990}$$

$$r_s = 1 - 0.267$$

$$r_s = 0.733$$

iii) Between 1st judge and 3rd judge

$$r_s = 1 - \frac{6\sum(d)^2}{N^3 - N}$$

$$r_s = 1 - \frac{6 \times 156}{(10^3) - 10}$$

$$r_s = 1 - \frac{936}{1000 - 10}$$

$$r_s = 1 - \frac{936}{990}$$

$$r_s = 1 - 0.945$$

$$r_s = 0.0545$$

Since co-efficient of correlation is maximum in the judgement of 2nd and 3rd judges. It is to be concluded that they have the nearest approach to common tastes in beauty.

23. Calculate the coefficient of correlation from the following data by the spearman's Rank Differences method :

Prices of Tea (`)	75	88	95	70	60	80	81	50
Prices of Coffee(`)	120	134	150	55	110	140	142	100

Ans :

Price of Tea (X)	R_1	Price of Coffee (Y)	R_2	$R_1 - R_2 = d$	d^2
75	4	120	3	1	1
88	7	134	4	3	9
95	8	150	7	1	1
70	3	55	1	–2	4
60	2	110	2	0	0
80	5	140	5	0	0
81	6	142	6	0	0
50	1	100	8	–7	49
					64

$$r_s = 1 - \frac{6\sum(d)^2}{(N)^3 - N} = 1 - \frac{6 \times 64}{(8)^3 - 8}$$

$$r_s = 1 - \frac{384}{512 - 8} = 1 - \frac{384}{504}$$

$$r_s = 1 - 0.762$$

$$r_s = 0.238$$

24. Compute Rank coefficient of correlation for the following data :

X	15	20	28	12	40	60	20	80
Y	40	30	50	30	20	10	30	60

Ans :

X	R_1	Y	R_2	$R_1 - R_2 = d$	d^2
15	2	40	6	–4	16
20	3.5	30	4	–0.5	0.25
28	5	50	7	–2	4
12	1	30	4	–3	9
40	6	20	2	4	16
60	7	10	1	6	36
20	3.5	30	4	–0.5	0.25
80	8	60	8	0	0
					81.5

$$r_s = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right]}{(N)^3 - N}$$

$$r_s = 1 - \frac{6\left[81.5 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^3 - 3)\right]}{(8)^3 - 8}$$

$$r_s = 1 - \frac{6[81.5 + 0.5 + 2]}{512 - 8}$$

$$r_s = 1 - \frac{6(84)}{504}$$

$$r_s = 1 - \frac{504}{504}$$

$$r_s = 1 - 1$$

$$= 0$$

25. Find the coefficient of rank correlation between the marks obtained in Mathematic (X) and those in Statistics (Y) by 10 students of certain class out of a total of 50 marks in each subject.

Students No.	1	2	3	4	5	6	7	8	9	10
X	12	18	32	18	25	24	25	40	38	22
Y	16	15	28	16	24	22	28	36	34	19

Ans :

X	R_1	Y	R_2	$R_1 - R_2 = d$	d^2
12	1	16	1.5	–0.5	0.25
18	1.5	15	1	0.5	0.25
32	8	28	7.5	0.5	0.25
18	1.5	16	1.5	0	0
25	6.5	24	6	0.5	0.25
24	5	22	5	0	0
25	6.5	28	7.5	1	1
40	10	36	10	0	0
38	9	34	9	0	0
22	4	19	4	0	0

$$r_s = 1 - \frac{6\left[\sum d^2\ \tfrac{1}{12}(m^3 - m) + \tfrac{1}{12}(m^3 - m)\right]}{N^3 - N}$$

$$r_s = 1 - \frac{6\left[2 \times \tfrac{1}{12}(4^3 - 4) + \tfrac{1}{12}(4^3 - 4)\right]}{(10)^3 - 10}$$

$$r_s = 1 - \frac{6\left[2 \times \tfrac{1}{12}(64 - 4) + \tfrac{1}{12}(64 - 4)\right]}{1000 - 10}$$

$$r_s = 1 - \frac{6[2 \times 5 \times 5]}{990}$$

$$r_s = 1 - \frac{300}{990}$$

$$r_s = 1 - 0.303$$

$$r_s = 0.697$$

4.5 Regression Analysis

Q22. Define the term Regression. Explain the utility of Regression Analysis.

Ans : **(Imp.)**

The dictionary meaning of the term ‘regression’ is the act of the returning or going back. The term ‘regression’ was first used by Sir Francis Galton in 1877 while studying the relationship between the heights of father and sons. Regression analysis is a technique used for the modeling and analysis of numerical data consisting of values of a dependent variable (response variable) and of one or more independent variables.

1. **Dependant Variable** is the single variable being explained/ predicted by the regression model (response variable).

2. **Independent Variable** is the explanatory variable(s) used to predict the dependant variable (Predictor variable).

Definitions

"Regression is the measure of the average relationship between two or more variables in terms of the original units of data."

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

According to Taro Yamane, "One of the most frequently used techniques in economics and business research, to find a relation between two or more variable that are related causally, is regression analysis."

Utility of Regression Analysis

1. Regression analysis helps in establishing a functional relationship between two or more variables. Once this is established it can be used for various advanced analytical purposes.

2. Since most of the problems of economic analysis are based on cause and effect relationship, the regression analysis is a highly valuable tool in economics and business research.

3. This can be used for prediction or estimation of future production, prices, sales, investments, income, profits and population which are indispensable for efficient planning of an economy and are of paramount importance to a businessman or an economist.

4. Regression analysis is widely used in statistical estimation of demand curves, supply curves, production functions, cost functions, consumption functions, etc. Economists have discovered many types of production functions by fitting regression lines to input and output data.

Q23. What are the objectives of Regression Analysis.

Ans :

1. The first objective of regression analysis is to provide estimates of values of the dependent variable from values of independent variable. This is done with the help of the regression line. The regression line describes the average relationship existing between X and Y variables, more precisely, it is a line which displays mean values of Y for given values of X.

2. The second objective of regression analysis is to obtain a measure of the error involved in using the regression line as a basis for estimation. For this purpose standard error of estimate is obtained. This helps in understanding the correlation existing between X and Y.

3. In general, we can model the expected value of y as an n^{th} order polynomial, yielding the general polynomial regression model

$$Y = a_0 + a_1x + a_2x^2 + a_2x^2 + a_3x^3 + \ldots + a_nx^n + \varepsilon$$

Conveniently, these models are all linear from the point of view of estimation, since the regression function is linear in terms of the unknown parameters ao,ai,.... Therefore, for least squares analysis, the computational and inferential problems of polynomial regression can be completely addressed using the techniques of multiple regressions. This is done by treating x, x^2, ... as being distinct independent variables in a multiple regression model.

Q24. What are the assumptions of Regression Analysis.

Ans :

The following assumptions are made while making use of the regression technique:

1. There exists an actual relationship between the dependent and independent variables.

2. The regression analysis is used to estimate the values within the range for which it is valid and not for the values outside its range.

3. The relationship that existed between the dependent and independent variables remains the same till the regression equation is calculated.

4. The dependent variable takes any random value but the values of the independent variables are fixed quantities without error and are chosen by the analyst or the user.

5. In regression, we have only one .dependent variable in our estimating equation. However, we can use more than one independent variable.

Q25. What are the limitations of Regression Analysis.

Ans :

1. It assumes a linear relationship between two variables which need not be the case always.

2. It assumes a static relationship between the two variables over a period of time. However, relationships between variables can change with a change in other factors. For example, the change in demand for a given change in price can be estimated using regression. However, the impact of price on demand will be different when a family or a nation is poor and when such a family or nation has abundance of wealth or resources.

3. Regression analysis provides meaningful insights only up to a certain limit. For example, increasing production results in a decrease in marginal cost. However, beyond a certain point, increase in production can result in the costs going up.

Q26. Explain different types of Regression.

Ans :

The various types of Regression are as follows:

1. Simple Regression

In statistics, simple regression is the least squares estimator of a linear regression model with a single predictor variable. In other words, simple linear regression fits a straight line through the set of n points in such a way that makes the sum of squared residuals of the model (that is, vertical distances between the points of the data set and the fitted line) as small as possible.

2. Multiple Regression

Multiple regression analysis represents a logical extension of two-variable regression analysis. Instead of a single independent variable, two or more independent variables are used to estimate the values of a dependent variable. However, the fundamental concepts in the analysis remain the same.

For example, a college admissions officer wishing to predict the future grades of college applicants might use three variables (High School GPA, SAT and Quality of letters of recommendation) to predict college GPA. The applicants with the highest predicted college GPA would be admitted. The prediction method would be developed based on students already attending college and then used on subsequent classes. Predicted scores from multiple regression are linear combinations of the predictor variables. Therefore, the general form of a prediction equation from multiple regression is:

$$Y' = b, X, + b_2 X_2 + ... + b_k X_k + A$$

where Y' is the predicted score, *X,* is the score on the first predictor variable, X_2 is the score on the second, etc. The Y intercept is A. The regression coefficients (b_1, b_2, etc.) are analogous to the slope in simple regression.

3. Curvilinear Regression

The analysis of the linear regression model can be extended in a straightforward way to cover situations in which the dependent variable is affected by several controlled variables or in which it is affected non-linearly by one controlled variable.

For example, suppose that there are three controlled variables, x,, x_2 and x_3. A linear regression equation is of the form,

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

4. Polynomial Regression

Suppose that the dependent variable is a polynomial function of a single controlled variable. For example, in cubic regression, the regression equation is given by,

$$y = ao + ai\ x + a_2x^2 + a3X^3$$

This type of regression can be approached in the same way as multiple regressions. In the case of cubic regression we can substitute x, = x, x_2 = x: and x, = x^3. The least squares estimates of a_0, a,, a_2 and a_3 can then be obtained.

If the observations are taken in such a way that there are an equal number of observations on y at a series of equally spaced values of x, then it is computationally more efficient to use the method of orthogonal polynomials.

4.5.1 Linear and Multiple Regression Analysis

Q27. Explain brielfy about Least Square Fit of Linear Regression.

Ans : **(Imp.)**

The regression equation of Y on X is expressed as follows :

$$Y = a + bX$$

It may be noted that in this equation 'Y' is a dependent variable, i.e., its value depends on X. 'X' is independent variable, i.e.,-we can take a given value of X and compute the value of Y.

'a'is "Y-intercept" because its value is the point at which the-regression line crosses the Y-axis, that is, the vertical axis, 'b' is the "slope" of line. It represents change in Y variable for a unit change in X variable.

'a' and 'b' in the equation are called numerical constants because for any given straight line, their value does not change.

If the values of the constants 'a' and 'b' are obtained, the line is completely determined. But the question is how to obtain these values. The answer is provided by the method of Least Squares which states that the line should be drawn through the plotted points in such a manner that the sum of the squares of the deviations of the actual Y values from the computed Y values is the least, or in other words, in order to obtain line which fits the points best Z(Y - Y_c)$^:$, should be minimum. Such a line is known as the line of 'best fit'.

A straight line fitted by least squares has the following characteristics:

1. It gives the best fit to the data in the sense that it makes the sum of the squared deviations from the line, V(Y – Y_c) smaller than they would be from any other straight line. This property accounts for the name 'Least Squares'.
2. The deviations above the line equal those below the line, on the average. This means that the total of the positive and negative deviations is zero, or X(Y - Y_c) = 0.
3. The straight line goes through the overall mean of the data (x, y).
4. When the data represent a sample from a large population the least squares line is a 'best' estimate of the population regression line.

Regression Equations

The regression equations express the regression lines. As there are two regression lines, so there are two regression equations. The regression equation X on Y describes the variation in the values of X for the given changes in Y, and used for estimating the value of X for the given value of Y, Similarly, the regression equation Y on X describes the variation in the values of Y for the given changes in X, and is used for estimating the value of Y for the given value of X.

1. **Regression Equation of Y on X :** With a little algebra and differential calculus it can be shown that the following two equations, if solved simultaneously, will yield values of the parameters a and b such that the least squares requirement is fulfilled:

$$\Sigma Y = Na + b\Sigma X$$

$\Sigma XY = a\Sigma X + b\Sigma X^2$

These equations are usually called the normal equations. In the equations ΣX, ΣXY, ΣX^2 indicate totals which are computed from the observed pairs of values of two variables X and Y to which the least squares estimating line is to be fitted and N is the number of observed pairs of values.

2. **Regression Equation of X on Y :** The regression equation of X on Y is expressed as follows:

$X = a + bY$

To determine the values of a and b, the following two normal equations are to be solved simultaneously:

$\Sigma Y = Na + b\Sigma Y$

$\Sigma XY = a\Sigma Y + b\Sigma Y^2$

PROBLEMS

26. From the following data obtain the two regression equations :

X :	**6**	**2**	**10**	**4**	**8**
Y :	**9**	**11**	**5**	**8**	**7**

Sol :

X	Y	XY	X^2	Y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\Sigma X = 30$	$\Sigma Y = 40$	$\Sigma XY = 214$	$\Sigma X^2 = 220$	$\Sigma Y^2 = 340$

Table : Obtaining Regression Equations

Regression equation of Y on X, $Y = a + bX$

To determine the values of a and b the following two normal equations are to be solved.

$\Sigma Y = Na + b\Sigma Y$

$\Sigma XY = a\Sigma Y + b\Sigma Y^2$

Substituting the values

$40 = 5a + 30b$ ... (1)

$214 = 30a + 220b$ ... (2)

Multiplying equation (1) by 6, $240 = 30a + 180b$ ... (3)

$214 = 30a + 220b$ .... (4)

Deducing equation (4) from (3) – 40b = 26 or b = – 0.65

Substituting the value of b in equation (1)

$$40 = 5a + 30 (-0.65)$$

or $5a = 40 + 19.5 = 59.5$ or $a = 11.9$

Putting the values of a and b in the equation

Regression equation of Y on X

$$Y = 11.9 - 0.65 X$$

Regression equation of X on Y X = a + bY and the two normal equations are:

$$\Sigma X = Na + b\Sigma Y$$

$$\Sigma XY = a\Sigma Y + b\Sigma Y^2$$

Substituting the values

$$30 = 5a + 40b \quad ... (5)$$

$$214 = 40a + 340b \quad ... (6)$$

Multiplying equation (5) by 8:

$$240 = 40a + 320b \quad ... (7)$$

$$214 = 40a + 340b \quad ... (8)$$

From equation (7) and (8)

$$-20b = 26 \text{ or } b = -1.3$$

Substituting the value of b in equation (5);

$$30 = 5a + 40(-1.3)$$

$$5a = 30 + 52 = 82$$

$$\therefore \quad a = 16.4$$

Putting the value of a and b in the equation.

The regression equation of X of Y :

$$X = 16.4 - 1.3 Y.$$

27. From the following data, calculate the regression equations taking deviation of items from the mean of X and Y series.

X	6	2	10	4	8
Y	9	11	5	8	7

Sol :

Regression Equation of X on Y

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

$$\bar{X} = \frac{\Sigma X}{N} = \frac{30}{5} = 6$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{40}{5} = 8$$

$$\therefore \quad \bar{X} = 6, \quad \bar{Y} = 8$$

X	Y	$X - \bar{X}$ (x)	$Y - \bar{Y}$ (y)	$(X - \bar{X})^2 = x^2$	$(Y - \bar{Y})^2 = y^2$	xy
6	9	0	1	0	1	0
2	11	–4	3	16	9	–12
10	5	4	–3	16	9	–12
4	8	–2	0	4	0	0
8	7	2	–1	4	1	–2
$\Sigma X = 30$	$\Sigma Y = 40$	$\Sigma x = 0$	$\Sigma y = 0$	$\Sigma x^2 = 40$	$\Sigma y^2 = 20$	$\Sigma xy = 26$

Number of Pairs N = 5

Regression Co-efficient

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{-26}{20} = -1.3$$

$\therefore \quad X - \bar{X} = b_{xy}(Y - \bar{Y})$

$X - 6 = 1.3\,(Y - 8)$

$X - 6 = 1.3\,Y + 10.4$

$X = -1.3Y + 10.4 + 6$

$X = 1.3Y + 16.4$

(or)

$X = 16.4 - 1.3\,Y$

Regression Equation of Y on X

$$(Y - \bar{Y}) = b_{yx}(X - \bar{X})$$

Regression Coefficient

$$b_{yx} = \frac{\Sigma x^2}{\Sigma y^2} = \frac{-26}{40} = -0.65$$

$\therefore \quad Y - \bar{Y} = b_{yx}(X - \bar{X})$

$Y - 8 = -0.65\,(X - 6)$

$Y - 8 = -0.65x + 3.9$

$Y = -0.65x + 3.9 + 8$

$Y = -0.65x + 11.9$

or

$Y = 11.9 - 0.65x$

28. In trying to evaluate the effectiveness of its advertisement campaign, a firm compiled the following information.

Year	2007	2008	2009	2010	2011	2012	2013
Adv. Expenditure (Rs. in '000)	12	15	17	23	25	34	40
Sales Rs. in lakhs	5.0	5.6	5.8	7.0	7.2	8.8	9.2

Fit the regression line and estimate probable sales when Adv. expenditure is ` 60,000.

Sol :

Lines of regression formed using the principle of least square methods helps in estimating sales when expenditure is ` 60,000.

Let advertisement expenditure be ‘X’ and sales be ‘Y’

Regression equation of Y on X is,

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$\bar{X} = \frac{\Sigma x}{N} = \frac{166}{7} = 23.71 = 24$$

$$\bar{Y} = \bar{Y} = \frac{\Sigma y}{N} = \frac{48.6}{7} = 6.94 \cong 7$$

$\therefore \quad \bar{X} = 24, \ \bar{Y} = 7$

X	Y	X – 24 (x)	Y – 7 (y)	x^2	y^2	xy
12	5.0	–12	–2.0	144	4	24
15	5.6	–9	–1.4	81	1.96	12.6
17	5.8	–7	–1.2	49	1.44	8.4
23	7.0	–1	0	1	0	0
25	7.2	1	0.2	1	0.04	0.2
34	8.8	10	1.8	100	3.24	18
40	9.2	16	22	256	4.84	35.2
$\Sigma X = 166$	$\Sigma Y = 48.6$	$\Sigma x = -2$	$\Sigma y = -0.4$	$\Sigma x^2 = 632$	$\Sigma y^2 = 15.52$	$\Sigma xy = 98.4$

Since both the means are integers, the following formula will be used for computing b_{xy} and b_{yx}.

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2}; \ b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$$

$$\therefore \quad b_{yx} = \frac{98.4}{632} = 0.155;$$

$$b_{xy} = \frac{98.4}{15.52} = 6.34$$

$$y - \bar{y} = b_{yx}(X - \bar{X})$$

$$y - 7 = 0.155\,(X - 24)$$

$$y - 7 = 0.155\,X - 3.72$$

$$y = 0.155\,X - 3.72 + 7$$

$$\therefore \quad y = 0.155\,X + 3.28$$

Sales when advertisement expenditure (X) is ` 60,000

$$y = 0.155\,X + 3.28$$

$$= 0.155\,(60{,}000) + 3.28$$

$$= 9303.28 \text{ or } 9.30$$

Q28. Define multiple regression analysis. State its objectives, purpose and assumptions.

Ans :

Multiple Regression Analysis

Multiple regression analysis is an addition to the simple regression analysis/bivariate linear regression. It allows a metric dependent variable to get anticipated by multiple independent variables. A simple regression analysis predict only one dependent variable with one independent variable, multiple regression analysis is an analysis which explores the effect of multiple independent variables on a single, interval-scaled dependent variable. Prices, interest rates, seasonality, advertising intensity, consumer income and other economic factors are some of the possible independent variables. The expanded form of simple regression equation to represent a multiple regression analysis is as follows,

$$Y = \beta_0 + \beta_0 x_1 + \beta_2 x_2 + ... + \beta_n x_n + r_e$$

Where

Y = Expected value of dependent variable.

β_0 = Constant value ofy when the value of all independent variables is zero.

$\beta_0 ... \beta_n$ = Parameters of regression coefficient

x_n = K^{th} independent variable

r_e = Random Error

Objective/Purpose of Multiple Regression Analysis

The objective of multiple regression analysis is same as that of the simple regression analysis. As the objective of simple regression analysis is to measure the linear association between a dependent and an independent variable, in the same way, multiple regression is conducted with more number or multiple independent variables.

The following are the reasons for using multiple regression analysis,

1. To improve the anticipation of dependent variables, independent variables can better interprets the dependent variables only when one independent variable is considered.
2. To estimate the effect of the single independent variable on dependent variables and for each independent variables unstandardized coefficients are estimated.
3. To construct an equation in order to estimate the expected value of the dependent variable which can be anticipated by multiple independent variables.

Assumptions of Multiple Regression

The following are the assumptions of multiple regression model for reducing the probabilities of Type I and Type II errors,

1. The basis assumption of the multiple regression analysis is that there exists a linear relationship between dependent variable and multiple independent variables.
2. Exact standard regression can be analyzed accurately when there is a linear relationship between the dependent and independent variables or not.
3. It is assumed that the error term of dependent variable in multiple regression for specific values of independent variables normal distribution is followed i.e., for independent variables specific values are available and multiple values are available for dependent variable.
4. Dependent variable is normally distributed. For every set of specific values of independent variables, the dependent variable distribution

is normally distributed. The expected values of distribution will fall on the regression line. In the same way, specific values of independent variables for every set where equivalent dependent variables values follow normal distribution and vice versa.

5. It assumes that independent variables are measured by excluding the error term, as it does not have any relation with the error term in regression equation.

6. When variance of errors of dependent variables for each set of particular value of independent variable is same the assumption of homoscedasticity is followed.

7. It is assumed that if the error term's value for a specific observation is already known, then it is not possible to anticipate the value of next error ahead time.

8. Assumed as one error term is independent of other error term.

Q29. Explain briefly about:

(a) Regression coefficients in multiple regression

(b) R^2 in multiple regression.

Ans :

(a) Regression Coefficients in Multiple Regression

The traditional regression programs gives standardized parameter estimates such as β_1 and β_2 which can be taken as partial regression coefficients.

The partial correlation is the correlation between two variables Y and X_1, that controls the correlation of X_2 with Y. Therefore, a standardized regression model with only two independent variables are.

$$Y = \beta_1 x_1 + \beta_2 x_2 + e_r$$

β_1 and β_2 coefficients are partial regression coefficients, which shows the independent variable relationship that exists with dependent variable and also the other variable is related to the dependent variable. Partial regression coefficients represents moderate relationships until there is a correlation between independent variables.

The regression coefficients may not be authentic, when correlation between two independentyariables are high. The standardized regression coefficient (β) is used in a situation when the researchers want to know which independent variable will anticipate the dependent variable.

One great benefit of β is that it yields a constant scale or β's are directly comparable. Hence, the greater the definite standardized regression coefficient value, the more that specific independent variable is responsible to interpret the dependent variable.

(b) R^2 in Multiple Regression

In multiple regression, coefficient of multiple determination shows the percentage of variation in Y that interprets the combination of all independent variables.

Example

If the values of R^2 = 0.845, then it means that 84.5% of the variance in the dependent variable is described by the independent variable.

If two variables are not independent with each other, then R^2 for a multiple regression model is equal to the R^2 values of two different simple regression models. Ordinarily, the independent variables are commonly linked to one another which means that mode R^2 from a multiple regression models are less than the discrete R^2 values which results in individual regression models.

The R-squared analysis of the patter of residuals and hypothesis are used to assess a model fit where statistical significance is checked by an F-test of the overall fit, after which it conducts the t-test of individual parameters.

In multiple regression the coefficient of determination gives the measure of goodness of fit for the estimated regression model.

The Sum of Squares due to Errors (SSE) is the measure of the errors using the estimated regression equation in order to evaluate the value of dependent variable in the sample. It is represented in the equation as shown below,

$$SSE = (Y_i - \hat{Y}_i)^2$$

Sum of Squares due to Regression (SSR) is the measure that provides how much the regression model fits the data. It is represented in the equation as shown below,

$$SSR = \Sigma(\hat{Y}_i - \bar{Y})^2$$

The total Sum of Squares (SST) measures the total error involved in using the mean to evaluate the dependent variable which depicts the difference between the mean value of dependent variable and it's estimated value by the regression model. It is represented by the equation as shown below,

$$SST = \Sigma(Y_i - \bar{Y})^2$$

Thus, the relationship between SST, SSR and SSE is,

$$SST = SSR + SSE$$

The coefficient of determination is given as,

$$R^2 = \frac{SSR}{SST}$$

Q30. Discuss the statistical significance of multiple regression. What are the steps involved in interpreting a multiple regression?

Ans :

Statistical Significance in Multiple Regression

To test the statistical significance F-test is used. F-test allows to test the corresponding value of SSR (Sum of Squares Regression) and SSE (Sum of Squares Errors). The following formula is used to test the significance,

$$F = \frac{SSR / v}{SSE / n - v - 1} = \frac{MSR}{MSE}$$

Where,

SSR = Sum of Squares Regression

SSE = Sum of Squares Error

v = Number of Independent Variables

n = Number of Observations

MSR = Mean Squares Regression

MSE = Mean Squares Error

df (degree of freedom) are,

df Numerator = v

df Denominator = n – v – 1.

The standard error of the estimate can be calculated with the help of the following equation,

$$S_e = \left[\sum_{i=1}^{n} \frac{(y - \hat{y})^2}{n - p - 1}\right]^{1/2}$$

Where,

y = value of independent variable from regression equation

n = Number of Observations

p = Number of Predictors i.e., Independent variable Now consider the simple linear regression model as,

$$Y = \beta_0 + \beta_0 + \varepsilon$$

$$Y = \beta_0 + \beta_1 + \varepsilon$$

$\beta_1 \neq 0$ in order to have X and Y linearly related.

In multiple regression also t-test is used for this purpose to test the statistical significance of individual parameter. The following hypothesis are tested,

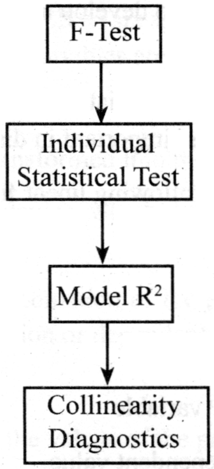
$$H_a : \beta_i \neq 0$$

$$H_0 : \beta_j \neq 0$$

$\beta_i \neq 0$ when H_0 is rejected when there exists a statistical relationship between dependent variable and 'i[th]' independent variable. But, when H_0 is not rejected, we can say that statistical relationship does not exist between the variables sampling distribution properties i.e., b_i provide the basis for the hypothesis test.

Steps in Interpreting A Multi Regression Model

Multiple regression models are usually involved to test some proposed theoretical model. These models can be explained with the help of the following steps,



Step 1: F-Test

The first basic step is to examine the model F-test. The model is to be dismissed if the test results are not significant.

Stpe 2: Individual Statistical Test

The next foremost step is to examine the individual statistical tests for each parameters. This is done when an independent variable with appropriate results are considered as a significant exploratory variable. The model should be run again, if an independent variable is not significant.

Step 3: Model R^2

When the researcher is more interested in prediction of the absolute value of R^2 rather than its explanation then R^2 model is to be examined. In this model, regression is conducted for forecasting purpose and the variable are more significant in explaining the dependent variables where the cutoff values for the model R^2 are not accurate.

Step 4: Collinearity Diagnostics

In regression analysis, multi collinearity is referred to enhance how the interdependent variable in a model is inter related. Individual parameter estimates becomes difficult to interpret when the multicollinearity is too high. Thus, majority of the regression programs compute Variance Factors (VIF) for variable above 5.0 that suggests problem with multi collinearity.

4.6 Discriminant Analysis

Q31. What is Discriminant Analysis? Explain the objectives of discriminant analysis?

Ans : **(Imp.)**

A discriminant analysis enables the researcher to classify persons or objects into two or more categories. For example, consumers may be classified as heavy and light users. With the help of such a technique, it is possible to predict the categories or classed which are mutually exclusive in which individuals are likely to be included. In many cases the classification will be dichotomous such as users and Non users high and low, and so on.

In discriminant analysis, A scoring system is used on the basis of which an individual or objects is assigned a score. This, in its turn, forms the basis for classifying an individual in the most likely class or category.

Suppose an individual is 25 years of age, earns an annual income of Rs. 6000 and has undergone formal education for a period of 17 years. Each of these three variables is given a weight, indicating is relative importance symbolically

$$Y = b_1 (25) + b_2 (60{,}000) + b_3 (17)$$

Where Y is a dependent variable, say, in this case credit score or rating. A certain limit is fixed of the value of Y below which all values will be classified in Group I and all the others in Group II. The values of b_1 and b_2 and b_3 indicate their importance. The numerical value of Y can be transformed into the probability of the individual being credit worthy.

If may be noted that in the linear discriminant the 'b' coefficients are similar to the regression coefficients. However the main difference is that while the regression coefficients are used to predict the value of the dependent variable, the discriminant coefficients are used to classify correctly as many individuals or objects as possible.

Objectives of Discriminant Analysis

1. Development of discriminant functions or linear combination of the predictor or independent variables, which will best discriminate between the categories of the criterion or dependent variable (groups).

2. Examination of whether significant differences exist among the groups, in terms of the predictor variables.

3. Determination of which predictor variable contribute to most of the inter group differences.

4. Classification of groups based on the values of predictor variables.

5. Evaluation of the accuracy of classification.

Q32. Explain the Model of Discriminate Analysis.

Ans :

The discriminant analysis involves linear combinations of the following form :

$$D = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + ... + b_k X_k$$

Where

D = Discriminate score

b_s = Discriminant coefficients or weights

X_s = predictors or independent variables.

The coefficients, or weights (b) are estimated so that the groups differ as much as possible on the values of the discriminant function. This occurs when the ratio of between group sum of squares to within group sum of squares for the discriminant scores is at a maximum. Any other linear combination of the predictors will result in a smaller ratio.

Q33. What are the different types of discriminant analysis?

Ans :

1. Two group Discriminant Analysis

When dealing with the data the researcher may encounter a case where criterion variable is nominal and the predictor variable is interval scaled.

In such cases Discriminant analysis is used. This analysis involves dividing the data into two groups or only two categories are involved viz. Discriminating the brand loyal from that of non brand loyal on the basis of dynamograph profiles. Acceptance or non acceptance of a certain product on the basis of the readership of various magazines.

Objective of the Two Group Discriminant Analysis

1. Finding the linear composites of the predictor variables that enables the analyst to separate the groups by maximizing within group variation.

2. Establishing procedures to assign a new individual to a particular group based on his profiles.

3. Testing whether there is significant difference between the mean predictor variables of the two groups.

4. Determining which variable account most for difference between the groups in mean profile.

The above objectives are the bases for the two major purposes and procedures for conducting discriminant analysis. The first procedure, Discriminant predictive analysis used to optimize the predictive functions.

Discriminant classification analysis, uses the predictive functions obtained in the previous case to classify the data set, thereby validating the predictive function and sometimes is used to classify new sets of data into categories.

2. Multiple Discriminant analysis

It is used to compute more than one discriminant function i.e. to group the data into more than two categories. All the assumptions and objectives are similar to two group discriminant analysis.

Linear Discriminant Analysis

Linear discriminant analysis is typically used to identify the characteristics which can accurately discriminate between the respondents who fall in one category from those who fall in another category.

Q34. Explain the Applications of Discriminant Analysis.

Ans :

Discriminant analysis can be used in atleast five different types of applications.

1. **Classifying and Describing Subjects:** The technique may be used as a way of classifying and describing subjects in two or more respective relevant groups at the same time.

2. **Tool for Improving the Quality:** Political scientists use discriminant analysis as a tool for improving the quality of their predictions about which subject belongs with which group and why.

3. **Determine Descriptive Variables:** Researchers often use the technique to determine which descriptive variables have the greatest power to discriminate between two or more groups of people.

4. **Serves as Check for Diagnoses:** The method may be used as a post hoc test, in which it serves as a check for diagnoses or predictions made on the bases of other types of evaluations.

5. **Gauge How Far Apart the Groups are Located:** Discriminant analysis may be used to gauge how far apart the groups are located on a set of descriptive characteristics. In this application, which is somewhat similar to the discrimination test, the distance between groups is based upon the location of the central tendency values (called centroids) for each group in two-dimensional space established by computer- generated functions.

Q35. Explain the steps for conducting the Discriminant Analysis.

Ans :

Step (1): Problem Formulation

Discriminant analysis starts with the formulation stage where the problem is formulated based on objectives, criterion and independent variables. Identified criterion variables must contain at least two mutually exclusive and exhaustive events. When the dependent variables are in interval (ratio) scale they must be converted into various groups.

For example, in case of measuring the behaviour of customers towards the product, behaviour can be measured on an 8-point scale, which can be categorized into positive, (1, 2, 3), neutral (4) and negative behaviour (5, 6, 7).

In further steps, the sample which is selected is divided into two parts. One part is used in the analysis of discriminant function and it is called as "analysis sample" ("estimation sample"). Another group is used for validation purpose and called as "holdout sample" ("validation sample"). The division of total sample into two groups depends upon the nature of distribution in the total sample. For example, if a sample consists of 20% of consumers earning a salary of ` 10,000 p.m while 80% are earning a salary of ` 5000 p.m, then each of the estimation and holdout groups must be represented by the same percentage of consumers.

Thus, the division of the total sample into analysis and validation groups should be a continuous process.

Step (2): Coefficient Estimation

After identifying analysis and validation sample, the next step is to estimate the coefficients of discriminant function. For this two methods are available. They are,

(a) Direct discriminant analysis method

(b) Step-by-step discriminant analysis method.

(a) Direct Discriminant Analysis Method

In case of direct discriminant analysis method, discriminant function is estimated by simultaneously including all the predictors irrespective of their level of independence. The direct method is mostly suitable in situations where the theoretical models are used and where the researchers want to determine its discriminant function on each and every estimator.

(b) Step-by-Step Discriminant Analysis Method

In case of step-by-step discriminant analysis method, estimator variables enter the analysis process sequentially depending on their respective capability of discriminating other variables within the group. This method is best suitable in situations, where the researcher wants to include a subset of the estimator variables in the discriminant function. In case of two-group discriminant analysis, only one discriminant function is analyzed or estimated.

Step (3): Significance Determination

The interpretation of statistically insignificant discriminant functions has no meaning. Statistical tests like Wilks' λ and other tests are used to test the null hypothesis which states that the means of all discriminant functions of all the groups are equal. For example, λ test is used in estimating the significance level and F-test is used in the calculation of likelihood ratio. As in MINITAB no tests are conducted to find out the significance level, the results are interpreted directly after the rejection of null hypothesis.

Step (4): Results Interpretation

The method of interpretation of discriminant analysis and multiple regression analysis is the same. The coefficient values of one estimator depends upon the other estimators of discriminant function. Even though, the coefficient signs are upredictable they reveal the functional values of variables and link them to their respective groups.

As there are no definite measures to test the significance of an estimator against the other estimators of the group, one way to find out the significance is by examining the dominant position played by the coefficients of discriminant function. Usually, estimators with large standardized coefficients will have more discriminating power than smaller coefficients.

One more way to find out the significance of the estimator is by examining structure correlation or discriminant loading. The correlation existing between the estimator and discriminant function indicates the difference between the estimator and its functions. Higher correlation indicates higher significance level of a particular estimator.

Thus, before interpreting the results, assessment of their validity is essential.

Step (5): Validation

Computer programs like SPSS provide leave-one-out validation technique in which the discriminant model is estimated repeatedly (number of times) depending upon the size of the sample. In this technique, after identifying respondents, predictions are made for that respondent using discriminant model. As each estimation leaves out one respondent, it is very difficult to use this model in case of large samples. So, in order to make this task easier, the total sample is divided into analysis sample and holdout sample.

The weights projected in the analysis sample and variable values of the holdout sample are multiplied by one another in order to bring out scores for the different cases of holdout sample. Finally, these cases are allocated to the groups according to the discriminant scores and decision rule. Tor example, in case of two-group discriminant analysis a case is allocated to that group which has the nearest centroid.

4.7 Exploratory Factor Analysis

Q36. Define Exploratory Factor Analysis. What are the objectives of Exploratory Factor Analysis?

Ans : **(Jan.-20)**

Factor Analysis was first used by Charles spearman. Psychologists use it as a technique of indirect measurement. When the test human personality and intelligence, a set of questions and tests are developed for this purpose.

This approach is based on the assumption that the underlying structure in answering the questions would be the same in the case of different respondents.

Although it is in the field of psychology that factor analysis has its beginning, it has since been applied to problems in different area including marketing.

In regression analysis, the problem is to predict the value of a dependent variable on the basis of one or more independent variables.

Unlike the regression analysis, factor Analysis is not based on the usual distinction between dependent and interdependent variables, instead it rather considers all the variables simultaneously.

Objectives of Factor Analysis

First, it simplifies the data by reducing a large number of variables to set a small number of variables.

Secondly, It analyses the interdependence of interrelation ships among a total set of variables.

Factor Analysis Model

Mathematically, factor analysis is some what similar to multiple regression analysis, in that each variable is expressed as a linear combination of underlying factors.

The amount of variance a variable shares with all other variables included in the analysis is referred to as communality.

If the variables are standardized, the factor model may be represented as

$$X_i = A_{i1} F_1 + A_{i2} F_2 + A_{i3} F_3 + \ldots A_{im} F_m + V_i U_i$$

Where

X_i = I the standardized variable

A_{ii} = Standardized multiple regression coefficient of variable i on common factor i

F = Common factor i

V_i = Standardized regression coefficient of variable i on unique factor i

U_i = The unique factor for variable i

M = Number of common factors

The unique factors are uncorrelated with each other and with the common factors. The common factors themselves can be expressed as linear combinations of the observed variables

$$F_i = W_{i1} X_1 + W_{i2} X_2 + W_{i3} X_3 + \ldots + W_{ik} X_k$$

Where

F_i = estimate of i the factor

W_i = weight or factor score coefficience

K = Number of variables

It is possible to select weights or factor score coefficient so that the first factor explain the largest portion of the total variance. Then a second set of weights can be selected, so that the second factor accounts for most of the residual variance, subject to being uncorrelated with the first factor.

This same principle could be applied to selecting additional weights for the additional factors. Thus the factors can be estimated so that their factor scores, unlike the values of the original variables, are not correlated.

Q37. What are the advantages and disadvantages of factor analysis?

Ans : **(Imp.)**

Advantages of Factor Analysis

In marketing research, factor Analysis can be useful in several ways.

1. It can bring out the hidden or latent dimensions relevant in the relationship among product preferences.

   Sometimes, the product characteristics influencing the consumer preference are not clear. In this case factor analysis can be helpful by revealing more important characteristics of the product, underlying the relationships among product preferences.

2. Factor Analysis can also be used to find out certain relationships among observed values, which though they exist, are obscure.

3. It is extremely useful when a large mass of data is to be simplified and condensed.

4. It can be used in clustering of products or people.

5. It is used in Market segmentation, advertising studies, pricing studies and product research.

Disadvantages of Factor Analysis

1. The technique of factor Analysis does not necessarily lead to the discovery of fundamental or basic categories in a filed of investigation. Sometimes relevant factors may be leftout.

2. Factor Analysis is a complicated tool and should be used by the researcher only when he has a good understanding of the technique. It is only involves in large number of variables, say 50 and it costly.

3. It is the reliability of results sometimes questionable a Factor analysis carried out form one half of the data might give quite different results from these obtained from the second half of the data..

4. Yet another limitation of technique is that its utility depends to a large extent on the judgement of the researcher. He has to make a number of decision as to how the factor analysis will come out.

5. Factor Analysis is unable to give unique solution result.

6. Thurstone mentions the use of factor analysis should not be made where fundamental and fruitful concepts one well formulated and tested. It may be used especially in those dominations where basic and fruitful concepts are essential lacking and where crucial experiments have been difficult to conceive.

7. The factor Analysis can be used with advantage of in the case of exploratory research. Many time Factor analysis is used just because it exists, without examining its suitability.

Q38. Explain the various Methods of Factor Analysis.

Ans :

Factor analysis is a generic name given to a set of computational procedure for data reduction. Therefore, data summarization or reduction can be done with several methods. However, they do not necessarily give the same results. As such factor analysis is not a unique method but a set of techniques with different computational process. Some of the important methods for data summarization or reduction that come under the generic name factor analysis are as follows:

1. Centroid method
2. Principal components analysis
3. Maximum likelihood method
4. Unweighted least squares method
5. Generalized least squares method
6. Principal axis factoring method
7. Alpha factoring method
8. Image factoring method

There are different methods for factor analysis. These methods differ with respect to the algorithm or the steps to find factors. There are several statistical packages to solve factor analysis. However, the principal component analysis is the most widely used method for EFA, commonly known as factor analysis, and we would specifically be focusing on understanding the same.

Q39. Explain the statistics techniques associated with factor analysis.

Ans :

The key statistics associated with factor analysis are as follows.

1. Correlation Matrix

A correlation matrix is a lower triangle matrix showing the simple correlations, r, between all possible pairs of variables included in the analysis. The diagonal elements, which are all 1, are usually omitted.

2. Communality

Communality is the amount of variance a variable shares with all the other variables being considered. This is also the proportion of variance explained by the common factors.

3. Eigen Value

The eigenvalue represents the total variance explained by each factor.

4. Factors Loading

Factor loading are simple correlations between the variables and the factors.

5. Factors Loading Plot

A factor loading plot is a plot of all the original variables using the factor loadings as coordinates.

6. Factor Matrix

A factor matrix contains the factor loading of all the variables on all the factors extracted.

7. Factors Scores

Factor scores are composite scores estimated for each respondent on the derived factors.

8. Percentage of Variance

This is true percentage of the total variance attributed to each factor.

9. Residuals

Residuals are the differences between the observed correlations, as given in the input correlation matrix and the reproduced correlations, as estimated from the factor matrix.

10. Scree Plot

A scree plot is a plot of the eigen values against the number of factors in order of extraction.

Q40. Explain the Steps Involved in Conducting Factor Analysis.

Ans :

Like any other analysis, the first step of factor analysis is to define the problem. Next, the variables of factor analysis are to be identified and then proceed to construct factor. A step-by-step procedure for factor analysis is given below:

Step 1 : Defining the Factor Analysis Problem

At the onset, the analyst has to formulate the factor analysis problem by identifying the objectives of factor analysis. Then, the researcher uses his research, theory and perception to define the variables of factor analysis. The variables identified are based on interval or ratio scale. The measurement of variables should necessarily be appropriate. The sample size should be large and appropriate.

Step 2 : Constructing Correlation Matrix

Next, the ratings data obtained on the defined variables are used to construct a correlation matrix. A correlation matrix gives an insight to the interrelationship of the visible at a glance. It involve analytic procedures to construct a matrix of correlation between the variables. It is important that the variables are highly correlated for an appropriate factor analysis. Appropriateness of factor model is tested using statistics like Barlett's test of sphericity and KMO Measure of Sampling Adequacy.

Under Barlett's Sphericity test, determines of the correlation matrix are transformed using chi square statistic. If the test statistic value is large, the null hypothesis that the variables are not correlated is rejected. Thus, if null hypothesis cannot be rejected, it means the factor model is inappropriate.

KMO measure of sampling is an index that compares the extent of observed correlation coefficients with that of the partial correlation coefficients. Higher KMO value (generally greater than 0.5) implies that the pairs of variable are highly correlated and hence the factor model is appropriate.

Step 3 : Deciding the Methods of Factor Analysis to be Used

Knowing that the factor model is appropriate, the analysts have to know decide on the right method of factor analysts. An analysts approach to obtain factor score coefficients influences the method of factor analysis to be used. Different intricate methods like unweighted least squares, generalized least squares, maximum likelihood, alpha method and image factoring can be used for factoring. Alternatively, two basic and popular methods used for factor analysis are principal components and common factor analysis.

Step 4 : Determining how Many Factors are Required

Factor analysis aims at summarizing data. So, the number of factors should be small, while retaining the maximum information from the original variables.

The analysts can use prior knowledge to determine the number of factors. It is called a Priori determination procedure for determining the number of factors. Different other procedure used to determine the ideal number of factors are mentioned below :

1. **Eigen Values :** Under this method, eigenvalues of the variables that signify the amount of variance associated with the factor determine the number of factors. Only variables with eigenvalues greater than 1 are considered ignoring the rest.

2. **Scree Plot :** Under scree plot method, the shape of the plot determines the number of factors to be extracted. The plot is characterized by a steep slope of factors with large eigen values. There is a break between this slope and a trailing off associated with the other factors. The break at the trailing of point is called scree. The point at which the scree starts is the number of factors to be obtained to gain maximum variance. It gives more or less same number as that given by eigenvalues method.

3. **Percentage of Variance :** Another statistic used for determining the number of factors is percentage of variance. When the

cumulative percentage of variance reaches a satisfactory level, usually 60% of the variance, the extraction of factors ceases.

4. **Split Half Reliability :** In this method, the sample is randomly divided into two equal halves. Factor analysis is constructed for both the samples and the factors with higher factor loading correspondence are selected.

5. **Significance Test :** Here, statistical significance is computed for the eigenvalues and the factors that are statistically significant are retained.

Step 5 : Rotating the Factors

Factor matrix obtained from the factor analysis consists of the coefficients that represent the correlation between the factors and the variables. The factors in the matrix are correlated with more than one variable and hence are difficult to interpret. For better interpretation the initial factor matrix should be rotated to obtain a simplified factor matrix.

While the communalities and the percentage of total variance described, remains same, the percentage of variance that each factor accounts for does change in rotation.

Different factors can be identified by following the three different methods of rotation. They are,

1. Orthogonal rotation
2. Varimax rotation and
3. Oblique rotation

1. Orthogonal Rotation

If the axis are perpendicular to each other, then the rotation is said to be orthogonal rotation. If the factors rotates along the axis which are right angles to each other then, orthogonal rotation is seen.

2. Varimax Rotation

It is most commonly used method of rotation. It is applicable only for uncorrected factors. Which minimizes the total number of variables having high loading factors, so that the interpretability of the factors should be maximum.

3. Oblique Rotation

The rotation when the axis aren't perpendicular to each other is called as oblique rotation. It is applicable to the perpendicular to the population having correlated factors. It plays an important role in simplifying the factor pattern matrix.

Step 6 : Interpreting the Factors

Once the factors are rotated they can be interpreted. Factor loadings are used to interpret to interpret factors. Factors can be interpreted based on the variables that have large factor loadings. Variables can be plotted using factor loadings as coordinates.

Variables at the end of an axis indicate high loadings on the same factor and describe that fact. Variables with small loadings on both the factors appear near origin. Variables that are near neither X-axis nor Y-axis are related to both the factors. Factors that cannot be expressed in terms of original variables are named as undefined factors.

Step 7 : Compute Factor Scores

After interpretation, analysts typically calculate factor scores. This helps in further reducing the original set of variables of subsequent analysis.

Factor is a linear combination of the actual variables. So, the factor scores can be computed as sum of products of the factors and the respective variables. Factor scores can be computed using computer programmes. These factor scores can be used for further analysis.

- **Select Surrogate Variables :** If factor scores are not computed analysts can select surrogate variables. Surrogate variables are substitute variables in which few of the original variables are pulled out for further analysis. By doing this, analysts can infer results in terms of the original variables rather than factor scores. The variables with high factor loadings on the factor matrix are taken as a surrogate variable for the associated factor.

Step 8 : Determine the Model Fit

Lastly, a model fit is determined. Factor analysis assumes that the observed correlation between the variables is due to the common factors.

This the correlations can be constructed from the estimated correlations between the variables and the factors.

The differences between the two correlations called residuals can be interpreted to determine the model fit. If large residuals are more, say more than five, it implies that the model is not fit and should be reconsidered.

Q41. What are the Applications of Factor Analysis?

Ans :

Applications of factor analysis are as follows,

- Factor analysis can be used for identifying market segments.
- Factor analysis can be applied to product research to determine the brand or product attributes influencing market demand.
- Factor analysis can also be applied to test certain hypotheses about the structure of a data set.
- It applies to advertising in understanding the media popularity. For instance, businessmen find less time for T.V. or movies but regularly use newspaper and Internet.
- It is also applicable to pricing studies for identifying the unique features of price--sensitive consumers.

4.8 Introduction to Structural Equation Modeling

Q42. What is Structural Equation Modeling? (SEM)

Ans : **(Imp.)**

Structural Equation Modeling (SEM) signifies a framework for the covariances between observed variables, thus it is also called as 'Covariance Structure Modeling'. SEM is found to be very useful to the researchers in testing hypothesis about the dimensionality of latent and observed variables and the relationship among them. SEM has become an alternative lo other multivariate methods that represent only a single relationship between the dependent and independent variables. Often, researchers refer to this model as LISREL (Linear Structural Relations) models.

Advantages of Structural Equation Modeling

The various advantages of SEM are as follows,

1. SEM facilitates in the simultaneous estimation of multiple and interrelated dependence relationships.
2. It can demonstrate the unobserved concepts/ latent variables in these relationship and also represents measurement error in the estimation process.
3. It helps in the modeling of complex multivariate relations which cannot be easily implemented elsewhere.
4. While regression analysis enables the modeling of observed variable, the factor analysis can model only unobserved variables. In contrast to these two analyses, the SEM include both observed and unobserved (latent) variables.
5. In order to get the unbiased estimates of the relationship between variables, the SEM provides an explicit modeling of measurement error. This helps the researchers in removing the measurement error from the estimates of regression/ correlation.

Q43. Explain Basic Terminologies used in Structural Equation Modeling (SEM).

Ans :

1. Manifest Variables

These are observed variables or indicators that the researcher has measured in the process of data collection. The researcher uses a suitable method or a scale to measure these variables. Some of these variables are called as predictors and outcome variables. In the SEM diagram, these variables are drawn as squares or rectangles.

2. Latent Variables

Latent variables are the un-observable variables in the model. These variables are hypothesized variables often known as constructs. These are derived from the manifest or measured variables in the measurement model. These are drawn as circles or ellipses in the SEM diagram.

3. Error

The error in the SEM is the variance left over after prediction of a measured variable. Often, it is denoted by the letter E.

4. Disturbance

It is the variance left over after prediction of a criterion or outcome variable and denoted by the letter D.

5. Exogenous or Independent Variable

The word exogenous is derived from Greek word exo, meaning being of external origin. These variables predict other variables or predictors. **Endogenous or dependent variable** The word endogenous is derived from Greek word endos, meaning being of internal origin. The endogenous variables are those variables that could be predicted by another variable or a set of variables. However, it may be mentioned that a predicted variable is endogenous even if it in turn predicts another variable and can become an exogenous variable.

6. Measurement Model

It is a specification of the relationship and is a part of the specified model that relates indicators to latent variables. The measurement model is the factor analytic part of SEM.

7. Structural Model

It is the part of the model that relates variable or factors to one another (as in regression and prediction), that is, this is the regression part of the SEM.

8. Direct Effect

Direct effect is the effect with respect to the regression coefficients for prediction.

9. Indirect Effect

It is the mediating effect of a variable on another variable through a third variable. For example, the effect of Xon *Y*through D.

10. Confirmatory Factor Analysis

This is an extension of the EFA. Confirmatory factor analysis (CFA) provides quantitative measures of the reliability and validity of the constructs used in the model. These constructs may be the outcome of the EFA.

11. Covariance Structure

This is the relationships between the variables based on their variances and covariances.

12. Mean Structure

These are the means (intercepts) estimated in the model.

Q44. What are the various components and statistical tools used in Structural Equation Modeling.

Ans :

Components of SEM

SEM is estimated in different stages. Most importantly, SEM consists of two major components (i) Measurement model (CFA) and (ii) Structural model (path analysis).

1. Measurement model (CFA)

It is a visual representation of the constructs, indicator variables, and interrelationships between the variables of the specified model. Confirmatory factor analysis (CFA) provides quantitative measures of the reliability and validity of the constructs used in the model.

2. Structural model (path analysis)

It is a set of dependency relationships that links constructs of the hypothesized model. It determines whether relationships exist between the constructs. Along with the statistics in CFA, it helps in accepting or rejecting the hypothesis.

Statistical Tests in SEM

The primary objective of SEM is to statistically test the interrelationship between the variables in simultaneous relationships. The following are different statistics and tests relating to the estimation of the SEM.

1. Goodness of fit is used to find how much is the estimated covariance matrix equal to observed covariance matrix.
2. Validity and reliability of measurement model are tested with the relevant statistics.

3. Significant and meaningfulness of structural relationship between the variables are tested. There are a few rules of thumb to judge the statistical significance relating to the interrelationships.

4.9 Cluster Analysis

Q45. Define Cluster Analysis. Explain the features of Cluster Analysis.

Ans : **(Aug.-21, Imp.)**

Cluster Analysis is used to classify persons or objects into a small number of mutually exclusive and exhaustive groups. There should be high internal (within - Cluster) homogeneity and high external (between cluster) heterogeneity.

In marketing research, cluster Analysis has been increasingly used because of its utility in resolving the problem of classifying consumers, products etc., at the end of our discussion on cluster analysis, its several uses in marketing will be specified.

Features Of Cluster Analysis

Let us look at some of the features of cluster analysis:

1. Cluster variate in cluster analysis refers to set of variables used to make a cluster or compare entities on the basis of their similarities and dissimilarities.
2. It is one of the multivariate techniques that do not estimate the variate like in the case of EFA empirically, but uses the predetermined variables specified by the re-searcher.
3. Accordingly, it may be observed that cluster analysis differs from EFA as the former groups objects, whereas the later groups variables.
4. Cluster analysis is an exploratory technique that is mosdy descriptive. We cannot use a statistical test to generalize the findings of a sample to the population. Therefore, it is said that cluster analysis is non-inferential. However, if the sample is a representative one, the results or findings of cluster analysis of a sample can be extended to the population.
5. One of the most important characteristics of cluster analysis is that the solutions are not unique. The output of the cluster analysis may be different and thus the cluster membership may change when different sets of clustering variables are used. More-over, addition and deletion of cluster variates may have significant impact on the output, that is, group composition or group membership.

In a hierarchical cluster analysis, the Euclidian distance in the dendrogram is used to find clusters. It is done manually and several trial and error methods are applied. In the dendrogram there is a horizontal line with a scale of 0-25, which is known as Euclidia distance. It is also termed as rescaled distance cluster combine. The more the distance, the more the dissimilarity in the clusters an vice versa.

Q46. What are the Objectives of Cluster Analysis?

Ans :

Cluster analysis is widely used in market segmentation, data reduction and data classification. Following are the objectives of cluster analysis.

1. Aid in Market Segmentation

Clustering of consumers of similar liking and needs, economic class and buying behaviours.

2. Identifying Consumer Behaviours

Clustering of customers depending on their buying behaviour, attitudes, perceptions etc.

3. Identifying Test Markets : Cluster of cities or locations for test marketing.

4. Classifying and Reducing Marketing Data: Automating or systematic clustering of marketing research data.

Q47. Explain the Methods of Cluster Analysis.

Ans :

The procedure of clustering can be Hierarchical and Non-hierarchical (K-means clustering). It is shown in the figure Below.

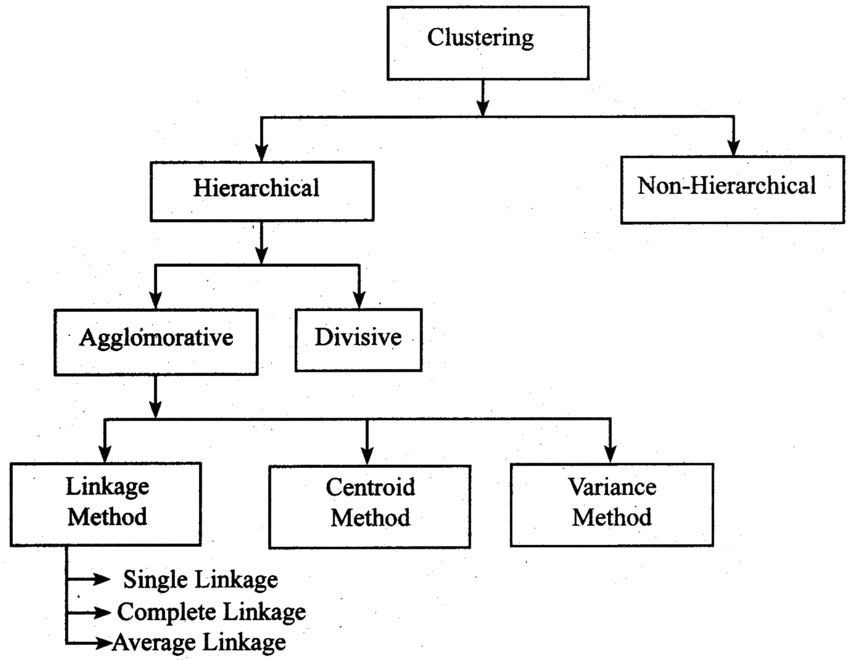


Fig.: Cluster Analysis Procedure

1. Hierarchical

The hierarchical method create dendrogram i.e., tree like structure and is further classified into,

(a) Agglomerative

(b) Divisive.

(a) Agglomerative

It begins as an independent cluster and ends as a single cluster, at every stage similar kind of clusters are combined. It is a most commonly used method of cluster analysis and is further classified as follows,

(i) Linkage Method

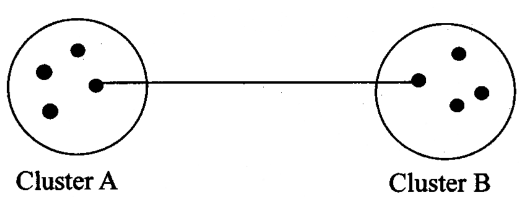
(ii) Centroid Method

(iii) Variance Method.

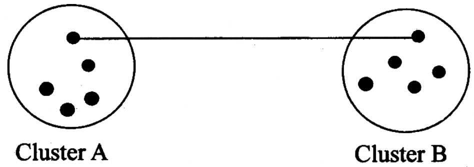
(i) Linkage Method

This method measures the distance. It further includes three linkages namely,

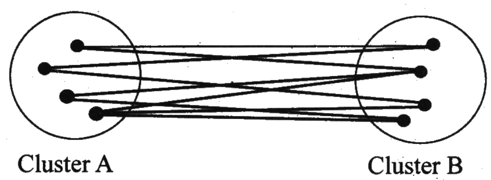
- **Single Linkage:** The clusters formed in the single linkage consists of minimum distance. So, it is also called as nearest neighbourhood rule.



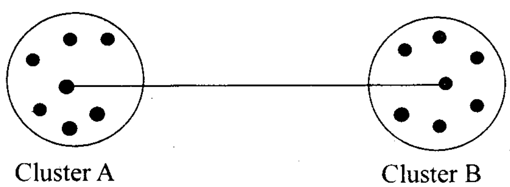
- **Complete Linkage:** The complete linkage method considers the maximum distance between two observations. So, it is also called as further neighbourhood rule.



- **Average Linkage:** The average linkage method considers the average distance between two observations.



- **Centroid Method:** This method take into consideration the distance between two centroids as it is the mean for all variables.



- **Variance Method:** The variance method is also known as ward's method. This method includes the squared distance from the means.

(b) Divisive

Divisive starts in a single cluster with all cases and ends after separating all clusters. Divisive divides all the clusters based on the differences between the cases with which it get started.

2. Non-Hierarchical (K-means Clustering)

The non-hierarchical method is also termed as means clustering. It is clear from the above stated methods that different set of clusters provide different results. The selection method is done on the basis of cluster clarity. Cluster analysis is a trial and error prodfess in case of hierarchical method. There are no particular tests which provides the validity of results rather than ANOVA.

Q48. Explain the steps involved in Conducting the Cluster Analysis.

Ans :

1. Formulate the Problem

The most important part of formulating problem is selecting the variables on which the clustering is based. The variables should be selected based on past research, theory or a consideration of the hypotheses being tested. In exploratory research the researcher should exercise judgement and intuition.

2. Select a Distance Measure

Because of objectives of clustering is to group similar objects together, some measures is needed to assess how similar or different the objects are. The most common approach is to measure similarity in terms of distance between pairs of objects. Objects with smaller distances between them are more similar to each other than are those at larger distances.

3. Select a Clustering Procedure

Clustering procedure can be hierarchical or non-hierarchical. Hierarchical clustering is characterized by the development of the hierarchy or tree-like structure. Hierarchical methods can be agglomerative or divisive. In agglomerative clustering we can use different methods, these are

i) Linkage methods – single, complete, average linkage

ii) Variance methods – word's method.

iii) Centroid methods.

The second type of clustering procedure, the non-hierarchical method. This method frequently referred to as K-means clustering. This method included,

1. Sequential threshold
2. Parallel threshold
3. Optimizing partitioning

4. Decide on the Number of Clusters

A major issue in cluster analysis is deciding on the number of clusters. Although there are no hard and fast rules, some guidelines are available.

i) Theoretical, conceptual or practical considerations may suggest a certain number of clusters.

ii) In hierarchical clustering, the distances at which clusters are combined can be used as criteria. This information can be obtained from the agglomeration schedule or from the dendrogram.

iii) In non-hierarchical clustering, the ratio of total within group variance to between group variance can be plotted against the number of clusters.

iv) The relative size of the cluster should be meaningful.

5. Interpret and Profile the Clusters

Interpreting and profiling clusters involves examining the cluster centroids. The centroids represent the mean value of the objects contained in the cluster on each of the variables. The centroids enable us to describe each cluster by assigning it a name or label. If the clustering program does not print this information if may be obtained through discriminant analysis.

6. Assess Reliability and Validity

Given the several judgments entailed in cluster analysis, no clustering solution should be accepted without some assessment of its reliability and validity. Formal procedures for assessing reliability and validity of clustering solutions are complex land not fully defensible.

Q49. What are the statistics tools Associated with Cluster Analysis?

Ans :

The following statistics and concepts are associated with cluster analysis.

1. **Agglomeration schedule:** A agglomeration schedule gives information on the objects or cases being combined at each stage of a hierarchical clustering process.

2. **Cluster centroid :** The cluster centroid is the mean values of variables for all the cases or objects in a particular cluster.

3. **Cluster centers :** The cluster centers are the initial starting points in non-hierarchical clustering. Clusters are build around these centers or seeds.

4. **Cluster membership :** Cluster membership indicates the cluster to which each object or case belongs.

5. **Dendrogram :** A dendrogram or tree graph, is a graphical device for displaying clustering results. Vertical lines represent clusters that are joined together. The position of the line on the scale indicates the distance at which clusters were joined. The dendrogram is read from left to right.

6. **Distance between cluster centers :** These distance indicates now separated the individual pairs of clusters are clusters that are widely separated are distinct, and therefore desirable.

7. **Icicle diagram :** An icicle diagram is a graphical display of clustering result, so called because it resembles a row of icicles hanging from the eaves of a house. The columns correspond to the objects being clustered and the rows.

4.10 Conjoint Analysis

Q50. Define conjoint analysis. What are the objectives of conjoint analysis.

Ans : **(Imp.)**

Conjoint analysis attempts to determine the relative importance consumers attach to salient attributes and the utilities they attach to the levels of attributes. This information is derived from consumers' evaluations of brands, or brand profiles composed of these attributes and their levels. The respondents are presented with stimuli that consist of combinations of attribute levels. They are asked to evaluate these stimuli in terms of their desirability.

Conjoint procedures attempt to assign values to the levels of each attribute, so that the resulting values or utilities attached to the stimuli match, as closely as possible, the input evaluations provided by the respondents. The underlying assumption is that any set of stimuli, such as products, brands, or stores, is evaluated as a bundle of attributes.

Like multi-dimensional scaling, conjoint analysis relies on respondents' subjective evaluations. Conjoint analysis seeks to develop the part-worth utility functions describing the utility consumers attach to the levels of each attribute. The two techniques are complementary.

Objectives of Conjoint Analysis

1. Determine Relative Importance of Attributes

Determining the relative importance of attributes in the consumer choice process. A standard output from conjoint analysis consists of derived relative importance weights for all the attributes used to construct the stimuli used in the evaluation task. The relative importance weights indicate which attributes are important in influencing consumer choice.

2. Estimating Market Share of Brands

Estimating market share of brands that differ in attribute levels. The utilities derived from conjoint analysis can be used as input into a choice simulator to determine the share of choices and hence the market share, of different brands.

3 Determining Composition of Brand

Determining the composition of the most preferred brand. The brand features can be varied in terms of attribute levels and the corresponding utilities determined. The brand features that yield the highest utility indicate the composition of the most preferred brand.

4. Segmenting Market

Segmenting the market based on similarity of preferences for attribute levels. The part-worth functions derived for the attributes may be used as a basis for clustering respondents to arrive at homogeneous preference segments.

Q51. Explain the Statistics tools Associated with Conjoint Analysis.

Ans :

The important statistics and terms associated with conjoint analysis include :

1. Part-Worth Functions

The part-worth functions or utility functions describe the utility consumers attach to the levels of each attribute.

2. Relative Importance Weights

The relative importance weights are estimated and indicate which attributes are important in influencing consumer choice.

3. Attribute Levels

They denote the values assumed by the attributes.

4. Full Profiles

Full profiles or complete profiles of brands are constructed in terms of all the attributes by using the attribute levels specified by the design.

5. Pair-Wise Tables

In this the respondents evaluate two attributes at a time until all the required pairs of attributes have been evaluated.

6. Cyclical Designs

They are designs employed to reduce the number of paired comparisons.

7. Fractional Factorial Designs

They are designs employed to reduce the number of stimulus profiles to be evaluated in the full profile approach.

8. Orthogonal Arrays

They are a special class of fractional designs that enable the efficient estimation of all main effects.

9. Internal Validity

This involves correlations of the predicted evaluations for the hold-out or validation stimuli with those obtained from the respondents.

Q52. Explain the Procedure of Conjoint Analysis.

Ans :

Major steps in Conjoint Analysis

1. Formulate the Problem

In formulating the conjoint analysis problem, the researcher must identify the attributes and attribute levels to be used in constructing the stimuli. Attribute levels denote the value assumed by the attributed. From theoretical stand point the attributes selected should be silent in influencing consumer preference and choice.

2. Construct Stimuli

Two broad approaches are available for constructing conjoint analysis stimuli. The pair wise approach and the full profile procedure. In the pair wise approach also called two factor evaluations & the respondents evaluate two attributes as a time until all the possible pairs of attributes have been evaluated.

In pull full profile approach is also called multiple factor evaluations. Full or complete profiles of brands are constructed for all the attributes.

3. Decide on the form of Input Data

As in the case of MDS, Conjoint Analysis input data can be either non-metric or metric. For Nonmetric data, the respondents are typically required to provide rank order evaluations.

In the metric form, the respondents provide rating, rather then rankings. In this case Judgment the judgments are typically made independently. In conjoint analysis, the dependent variable is usually preference or intention to buy. In recent years, the use of rating has become increasingly common.

4. Select a Conjoint Analysis Procedure

The basic conjoint analysis model may be represented by the following formula.

Where

U(X) = Overall utility of an alternative

I_j = the part worth contribution or utility associated with the Itn level.

(j. j = 1, 2 – ki) of the attribute (i, i = 1,2 –m)

k_i = Number of levels of attributes

m = Number of attributes

x_{ij} = 1 If the jth of the ith attribute is present

The mathematical model expressing the formulated relationship between attributes and utility in conjoint analysis

5. Interpret the Result

The next is interpret the result to conduct the conjoint analysis. In this step, through functions to interpret the result.

6. Assessing Reliability & Validity

Several procedure are available for assessing be reliability and validity of conjoint analysis result.

i) The goodness of fit of the estimated model should be evaluated.

ii) Test-Retest reliability can be assessed by obtaining a few replaced judgments later in data collection.

iii) The evaluation for the bold our or validation stimuli can be predicted by the estimated part worth functions.

iv) If an aggregate analysis has been conducted the estimation sample can be split in several ways and conjoint analysis conducted on each sub sample. The result can be compared across sub samples to assess the stability of conjoint analysis solutions.

Q53. Explain the Applications of Conjoint Analysis.

Ans :

1. Optimum Product Design

Since all possible product concepts can be compared after adding their respective attribute levels part worth utilities, it is possible to determine the demand for different products out of any given set of available products in the marketplace. The demand levels can be converted into profit figures as cost of producing and marketing can also be calculated. These cost calculations are possible as the volume of operations and the features of the products are now known.

Thus, the optimum product can be chosen from the profits point of view (or any of the other given management's objective). Customer's differential rates of purchase of products are also duly considered at this stage. Quite often, a manager may like to know the effects of slight change in any of the attribute by his own company or the competitor's. Conjoint Analysis allows this kind of "What if' analysis very easily with the data base of part worth utilities. In fact, different kinds of scenarios can be simulated and the manager can optimise not only the product but other aspects of his marketing strategy. Similarly, whenever there is any change in competitor's actions or in the environment a fresh scenario can be-drawn for simulation. Of course, the simulation shall be limited to the attributes considered in the analysis. This feature does also help in increasing the shelf life of the conjoint analysis output.

2. Market Segmentation

Since the Conjoint Analysis is done at the individual customer level, the individual customer's identity can be retained throughout the analysis. Thus, consumers can be segmented according to their sensitivities to different product attributes. It is also possible to identify the customers segments which would be attracted most for the proposed product position. This helps in having a focussed matching between the chosen product position and the target customer segment. It can also help in identifying that part of competitor's market which needs to be. poached for snatching market share from them. Similarly, the same type of analysis can be done to identify the most vulnerable section of one's own market segment. Sometimes, an additional product offer appears to be quite attractive. But, this may be at the cost of cannibalisation. Conjoint Analysis can help in estimating the effects of cannibalisation as well. Thus, it helps in maximising net profits of the organisation.

3. SWOT Analysis

First of all, the part worth utility of the brand itself can tell about the relative brand strength. Similarly by looking at the other features of one's own and competitor's offers. Conjoint Analysis enables the marketers to conduct his detailed SWOT analysis.

4. Estimating Customer Level Brand Equity

Conjoint Analysis provides a bridge between the consumer level perceptions and the financial worth of the offers. This can be used for estimating the important parameter of brand equity at the consumer's level. There is scope of differentiating the "Loyal", "Acceptors" and "Switchers" for more accurate calculations of brand equity.

Q54. What are the Methods of Conjoint Analysis?

Ans :

Some of the important methods used for conjoint analysis are traditional method of conjoint analysis, choice-based method of conjoint analysis, and adaptive method of conjoint analysis.

1. Traditional Method

The earliest forms of conjoint analysis or the traditional method of conjoint analysis were known as the full profile studies. In this type of conjoint analysis, a small set of attributes (typically four to five) were used to create profiles and cards were used (or it could be with a schedule). The cards incorporate different combinations of attributes for rating by the respondents. These cards with the attributes are shown to the respondents to elicit their preferences of the card. Alternatively, a schedule can also be shown where different combinations have been listed in a sheet of paper. These combinations are rated by the respondents in a group by giving the preferences, for example, the first preference of the combination gets 1 and the second gets 2.'

2. Choice-based Method

In this method, a set of profiles is generated, which is known as stimuli. If there are, for example, three variables with three levels and three categories, there will be 3 x 3 × 3 = 27 combinations, which can be shown in the cards to the respondents. These are rated by respondents according to their preferences. Usually, less than six attributes are used in this approach. One of the differences is that the respondents chose a profile that is a combination rather than an individual factor as in the case of the traditional method to avoid complication. Therefore, the analysis will be more complicated once a large number of factors with different levels are issued. However, this method gives the respondent a choice with respect to interactions of factors through the combinations.

3. Adaptive Method

This method has been developed to accommodate a large number of factors that may go up to even 30. These are also known as the hybrid conjoint techniques. Like the choice-based method, this method employs a computerized process that gives different combinations in the form of an orthogonal design. Individual cards are generated and shown to respondents for putting up their choices. These cards are generated using computer software like SPSS.

Exercises Problems

1. Calculate Correlation for the data given below :

X	42	36	48	43	55	52	38
Y	132	120	140	143	142	148	122

[Ans : r = 0.856]

2. Calculate Co - efficient of Correlation and interpret the value :

Marks in QT	57	42	40	38	42	45	42	44	40	46	44	43
Marks in IOM	10	26	30	41	29	27	27	19	18	19	31	29

[Ans : r = –0.7285]

3. Calculate the coefficient of correlation and probable error from the following data :

X	1	2	3	4	5	6	7	8	9	10
Y	20	16	14	10	10	9	8	7	6	5

[Ans : r = –0.95; P.E. = 0.0208]

4. Fit a straight line regression equation of Yon X from the following data.

X	10	12	13	16	17	20	25	29
Y	10	12	24	27	29	33	37	42

[Ans : Y = 1.6, X – 1.65]

5. Find the two regression equations from the following data :

X	1	2	3	4	5
Y	2	3	5	4	6

[Ans : X -= 0.9 Y – 0.6;Y = 0.9 X + 1.3]

6. From the following data, obtain the two regression lines.

X	2	6	8	11	13	13	13	14
Y	8	6	10	12	12	14	14	20

[Ans : X=0.8125, Y + 0.25; Y = 0.8125 X +3.875]

7. From the data given below about the treatment of 250 patients suffering from a disease, state whether the new treatment is superior to the conventional treatment :

Treatment	Favourable	No. of patients Not favourable	Total
New	140	30	170
Conventional	60	20	80
Total	200	50	250

[Ans : Given for degree of freedom = 1, chi - square 5 per cent = 3.84)]

Short Question and Answers

1. What is ANOVA? What are its assumptions and applications?

Ans :

ANOVA

The variance test is also known as ANOVA. ANOVA is the acronym for Analysis of Variance. Analysis of variance is a statistical technique specially designed to test whether the means of more than two quantitative populations are equal i.e., to make inferences about whether those samples are drawn from the populations having the same mean.

The test is called 'F' test as it was developed by R.A Fisher in 1920's. The test is conducted in situations where we have three or more to consider, at a time an alternative procedure (to t-test) needed for testing the hypothesis that all samples could likely be drawn from the same population.

Example

Five fertilizers are applied to four plots, each of wheat and yield of wheat on these plots is given. We are interested in finding out whether the effects of these fertilizers on the yields are significantly different or, in other words, whether the samples have come from the same population. ANOVA answers this question.

2. Chi-Square Distribution.

Ans :

The square of a standard normal variable is called a chi-square (χ^2) variate with one degree of freedom (d.f)

$$\chi^2 = \left(\frac{x-\mu}{\sigma}\right)^2$$

Where 'χ' is a random variable following normal distribution with mean 'μ' is an standard deviationσ. If $\chi_1, \chi_2, \dots, \chi_v$ are independent random variables following normal distribution with means $\mu_1, \mu_2, \dots \mu_v$ and standard dartrous $\sigma_1, \sigma_2, \dots \sigma_v$ then,

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

χ^2 is the sum of square of V independent standard normal variates following chi-square distribution. χ^2 is a family of distributions one for each value of V.

Applications of χ^2-Distribution

Chi-square distribution has a number of applications,

1. Chi-square test of goodness of fit.
2. Chi-square test for independence of attributes.
3. To test if population has a specified value of variance.
4. To test equality of several population proportions.

3. Explain briefly about test for goodness of fit?

Ans :

One of the very popular applications of χ^2 test is test of goodness of fit. It enables us to ascertain how the theoretical distributions such as Binomial, Poisson, Normal etc. can fit into empirical distributions obtained from sample data. When an ideal frequency curve whether normal or some other type is fitted to the data, we are interested out how well this curve fits with the observed facts.

A test of the concordance (goodness of fit) of the two can be made just by inspection, but such a test is obviously inadequate. Precision can be secured by applying the χ^2 test. The following are the steps in testing the goodness of fit :

1. Null and alternative hypotheses are established, and a significance level is selected for rejection of the null hypothesis.
2. A random sample of independent observations is drawn from a relevant statistical population.

3. A set of expected or theoretical frequencies is derived under the assumptions that the null hypothesis is true. This generally takes the form of assuming that a particular probability distribution is applicable to the statistical population under consideration.

4. The observed frequencies are compared with the expected, or theoretical frequencies.

5. If the calculated value of χ^2 is less than the table value at a certain level of significance (generally 5% level) and for certain degrees of freedom the fit is considered to be good, i.e., the divergence between the actual and expected frequencies is attributed to random fluctuations of sampling.

4. Define Correlation. Explain the significance of Correlation.

Ans :

Meaning and Definition of Correlation

Correlation is the study of the linear relationship between two variables. When there is a relationship of 'quantitative measure between two set of variables, the appropriate statistical tool for measuring the relationship and expressing each in a precise way is known as correlation.

For example, there is a relationship between the heights and weights of persons, demand and prices of commodities etc.

Correlation analysis is the statistical tool we can use to describe the degree to which one variable is linearly related to another.

According to Croxton and Cowden, "The appropriate statistical tool for discovering and measuring the relationship of quantitative nature and expressing it in brief formula is known as correlation".

According to Tippet, "The effects of correlation are to reduce the range of uncertainty of our prediction".

The coefficient of correlation measures the degree of relationship between two set of figures. As the reliability of estimates depend upon the closeness of the relationship it is imperative that utmost care be taken while interpreting the value of coefficient of correlation, otherwise wrong conclusion can be drawn.

5. Define scatter diagram. Explain merits and demerits of scatter diagram.

Ans :

Scatter diagram is a special type of dot chart that is used for graphical representation of the relationship between two quantitative variables.

Merits of Scatter Diagram

1. Scatter diagram is a simple and attractive method of finding out the nature of correlation between two variables.

2. It is a non-mathematical method of studying correlation. It is easy to understand.

3. We can get a rough idea at a glance whether it is a positive or negative correlation.

4. It is not influenced by extreme items.

5. It is a first step in finding out the relationship between two variables.

Demerits of Scatter Diagram

1. The major limitation of the method is that it only gives a visual picture of the relationship of two variables. It only tells us whether there is correlation between the variables, and if so, then in which direction, positive or negative.

2. It does not give an idea about the precise degree of relationship as it is not amenable to mathematical treatment.

6. Merits and demerits of coefficient of correlation.

Ans :

Merits of Karl Pearson's Coefficient of Correlation

1. It takes into account all items of the variable(s).

2. It is a numerical measure and hence more objective.

3. It measures both direction as well as degree of change.

4. It facilitates comparisons between two series.

5. It is capable of further Algebraic treatment

6. It is more practical and hence popular and is more commonly used.

Demerits of Karl Pearson's Coefficient of Correlation

1. It is not easy to calculate as complex formulae are involved.
2. It is more time consuming compared to methods such as rank correlation
3. It assumes a linear relationship between the two variables which may not be correct
4. It is impacted by extreme values as it is based o mean and standard deviation.
5. It is not easy to interpret.

7. Features of Spearman's Correlation Coefficient

Ans :

- The sum of the differences of ranks between two variables shall be zero. Symbolically. Xd = 0.
- Spearman's correlation coefficient is distribution-free or non-parametric because no strict assumptions are made about the form of population from which sample observations are drawn.
- The Spearman's correlation coefficient is nothing but Karl Pearson's correlation coefficient between the ranks. Hence, it can be interpreted in the same manner as Pearsonian correlation coefficient.

8. Merits and demerits of rank correlation ?

Ans :

Merits

1. It is easy to calculate and understand as compared to Pearson's *r.*
2. When the ranks of different values of the variables are given, it is then the only method left to calculate the degree of correlation.
3. When actual values are given and we are interested in using this formula then, we have to give ranks to calculate correlation.
4. This method is employed usefully when the data is given in a qualitative nature like beauty, honesty, intelligence etc.

Demerits

1. This method cannot be employed in a grouped frequency distribution.
2. If the items exceed 30, it is then difficult to find out ranks and their differences.
3. This method lacks precision as compared to Pearson's co- efficient of correlation, as all the information concerning the variables is not used. It is just possible that the difference between r_k and r may be very insignificant.

9. Define the term Regression.

Ans :

The dictionary meaning of the term 'regression' is the act of the returning or going back. The term 'regression' was first used by Sir Francis Galton in 1877 while studying the relationship between the heights of father and sons. Regression analysis is a technique used for the modeling and analysis of numerical data consisting of values of a dependent variable (response variable) and of one or more independent variables.

1. **Dependant Variable** is the single variable being explained/ predicted by the regression model (response variable).
2. **Independent Variable** is the explanatory variable(s) used to predict the dependant variable (Predictor variable).

Definitions

"Regression is the measure of the average relationship between two or more variables in terms of the original units of data."

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

According to Taro Yamane, "One of the most frequently used techniques in economics and business research, to find a relation between two or more variable that are related causally, is regression analysis."

10. What are the limitations of Regression Analysis.

Ans :

1. It assumes a linear relationship between two variables which need not be the case always.
2. It assumes a static relationship between the two variables over a period of time. However, relationships between variables can change with a change in other factors. For example, the change in demand for a given change in price can be estimated using regression. However, the impact of price on demand will be different when a family or a nation is poor and when such a family or nation has abundance of wealth or resources.
3. Regression analysis provides meaningful insights only up to a certain limit. For example, increasing production results in a decrease in marginal cost. However, beyond a certain point, increase in production can result in the costs going up.

11. Multiple Regression Analysis

Ans :

Multiple regression analysis is an addition to the simple regression analysis/bivariate linear regression. It allows a metric dependent variable to get anticipated by multiple independent variables. A simple regression analysis predict only one dependent variable with one independent variable, multiple regression analysis is an analysis which explores the effect of multiple independent variables on a single, interval-scaled dependent variable. Prices, interest rates, seasonality, advertising intensity, consumer income and other economic factors are some of the possible independent variables. The expanded form of simple regression equation to represent a multiple regression analysis is as follows,

$$Y = \beta_0 + \beta_0 x_1 + \beta_2 x_2 + ... + \beta_n x_n + r_e$$

Where

Y = Expected value of dependent variable.

β_0 = Constant value ofy when the value of all independent variables is zero.

$\beta_0 ... \beta_n$ = Parameters of regression coefficient

x_n = K^{th} independent variable

r_e = Random Error.

12. Discriminant Analysis? Explain the objectives of discriminant analysis?

Ans :

A discriminant analysis enables the researcher to classify persons or objects into two or more categories. For example, consumers may be classified as heavy and light users. With the help of such a technique, it is possible to predict the categories or classed which are mutually exclusive in which individuals are likely to be included. In many cases the classification will be dichotomous such as users and Non users high and low, and so on.

In discriminant analysis, A scoring system is used on the basis of which an individual or objects is assigned a score. This, in its turn, forms the basis for classifying an individual in the most likely class or category.

Suppose an individual is 25 years of age, earns an annual income of Rs. 6000 and has undergone formal education for a period of 17 years. Each of these three variables is given a weight, indicating is relative importance symbolically

$$Y = b_1 (25) + b_2 (60,000) + b_3 (17)$$

Where Y is a dependent variable, say, in this case credit score or rating. A certain limit is fixed of the value of Y below which all values will be classified in Group I and all the others in Group II. The values of b_1 and b_2 and b_3 indicate their importance. The numerical value of Y can be transformed into the probability of the individual being credit worthy.

13. Advantages of Factor Analysis

Ans :

In marketing research, factor Analysis can be useful in several ways.

1. It can bring out the hidden or latent dimensions relevant in the relationship among product preferences.

   Sometimes, the product characteristics influencing the consumer preference are not clear. In this case factor analysis can be helpful by revealing more important characteristics of the product, underlying the relationships among product preferences.

2. Factor Analysis can also be used to find out certain relationships among observed values, which though they exist, are obscure.
3. It is extremely useful when a large mass of data is to be simplified and condensed.
4. It can be used in clustering of products or people.
5. It is used in Market segmentation, advertising studies, pricing studies and product research.

14. Advantages of Structural Equation Modeling

Ans :

The various advantages of SEM are as follows,

1. SEM facilitates in the simultaneous estimation of multiple and interrelated dependence relationships.
2. It can demonstrate the unobserved concepts/latent variables in these relationship and also represents measurement error in the estimation process.
3. It helps in the modeling of complex multivariate relations which cannot be easily implemented elsewhere.
4. While regression analysis enables the modeling of observed variable, the factor analysis can model only unobserved variables. In contrast to these two analyses, the SEM include both observed and unobserved (latent) variables.
5. In order to get the unbiased estimates of the relationship between variables, the SEM provides an explicit modeling of measurement error. This helps the researchers in removing the measurement error from the estimates of regression/ correlation.

15. Objectives of Cluster Analysis?

Ans :

Cluster analysis is widely used in market segmentation, data reduction and data classification. Following are the objectives of cluster analysis.

1. Aid in Market Segmentation

Clustering of consumers of similar liking and needs, economic class and buying behaviours.

2. Identifying Consumer Behaviours

Clustering of customers depending on their buying behaviour, attitudes, perceptions etc.

3. Identifying Test Markets : Cluster of cities or locations for test marketing.

4. Classifying and Reducing Marketing Data: Automating or systematic clustering of marketing research data.

UNIT V	**Time Series Analysis and Report Writing:** (a) Components, Models of Time Series–Additive, Multiplicative and Mixed models; Trend Analysis- Free hand curve, Semi averages, moving averages, Least Square methods and Index numbers – introduction, Characteristics and uses of index numbers, types of index numbers, unweighted price indexes, weighted price indexes, Tests of adequacy and consumer price indexes. (b) Importance of Report writing, Types of Research Reports, Report Preparation and presentation, Report structure, Report formulation, Guides for effective documentation, Research Briefings. Referencing styles and citation in Business Management Research.

5.1 Time Series Analysis

Q1. Define Time Series.

Ans :

An arrangement of statistical data in accordance with time of occurrence or in a chronological order is called a time series. The numerical data which we get at different points of time-the set of observations is known as time series.

Examples:

i) The Annual Production of Steel in India over the last 10 years;

ii) The Monthly Sales of a Chemical Industry for the last 6 months;

iii) The daily closing price of a share in the Calculate Stock Exchange;

iv) Hourly temperature recorded by the Meteorological office in a city;

v) Yearly Price or Quantity Index Numbers.

Mathematically, a Time Series is defined by the values Y_1, Y_2, —, Y_n of a variable Y at times t_1, t_2,....t_n

Here Y is a function of time t and Y, denotes the value of the variable Y at time t. The basic assumption is that changes witnessed over time in a sample group will be extrapolated to population.

i) According to Morris Hamburg, "A time series is a set of observations arranged in chronological order".

ii) According to Kenny and Keeping, "A set of data depending on the time is called time series".

iii) According to Croxton and Cowden, "A time series consists of data arranged chronologically".

iii) According to Patterson, "Time series consists of statistical data which are collected, recorded or observed over successive increments".

Q2. Explain the Utility of Time Series Analysis.

Ans : **(Imp.)**

i) It Helps in the Analysis of past Behaviour of a Variable

Analysis of past data discloses the effect of various factors on the variable under study. These studies isolate and analyze the effects of various sets of homogenous factors on the problem under study.

ii) It Helps in Forecasting :

The analysis of past condition is the basis of forecasting the future behaviour of the variable under study. For example, the analysis of a time series relating to income or wages or production would be the basis for forecasting future income or wages or production. This helps in making future plan of action.

iii) It helps in Evaluation of Current Achievement :

The review and evaluation of progress made on the basis of a plan are done on the basis of time series data. For example the progress of our Five-year plans is judged by the annual growth rates in the gross national product.

iv) It Helps in Making Comparative studies :

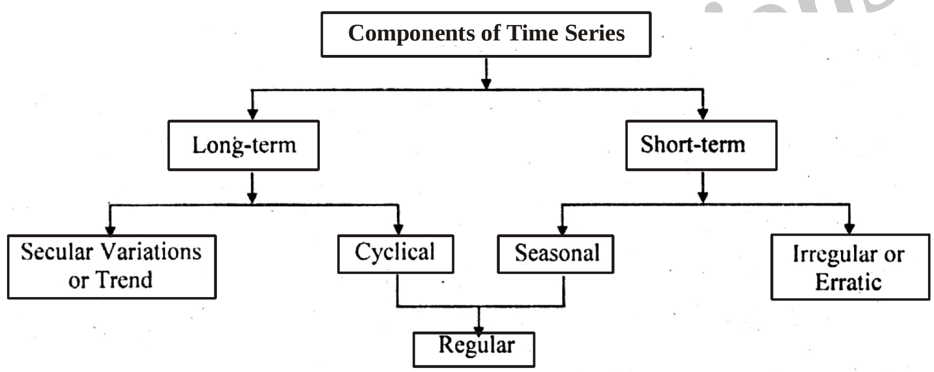
Once the data are arranged chronologically comparison between one time period and another is facilitated. It provides a scientific basis for making comparisons by studying and isolating the effects of various components of a time of series. It also helps in making regional comparison amongst data collected on the basis of time.

5.1.1 Components

Q3. What are the Components of Time Series ?

Ans : **(Nov.-20, Imp.)**

A lime series is a historical series of statistical data. Since these statistical data are historical, they are subject to the influences of all the changes that may occur over any period of time. The usual classification of the influences affecting statistical data recording over time is one based upon the nature of the influence. Classified in this manner, these influences are as follows :



1) **Secular Variations or Trend (T):** 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 simply trend.

   According to Simpson and Kafka: "Trend, also called secular or long-term trend, is the basic tendency of a series to grow or decline over a period of time. The concept of trend does not include short-range oscillations, but rather the steady movements over a long time".

   The trend thus refers to the general direction and the movement of time series considering a fairly long period of time. Most business and economic series are influenced by some secular forces of charge in which the underlying tendency is one of growth or decline, for example, in series concerning population, national income, agricultural production, currency in circulation, bank deposits, etc., an upward tendency is usually observed while a downward tendency is noticed in time series relating 10 birth and death rates specially in deaths by epidemics, tuberculosis, etc.

2) **Seasonal Variations (S) :** Seasonal variations refers to rhythmic forces of change inherent in most time series showing a regular or a periodic pattern of movement over a span of less than a year and has the same or almost the same pattern year after year.

   According to Hirsch "a recurrent pattern of charge within the period that results from the operation of forces connected with climate or custom at different times of the period."

Reasons for Seasonal Variations

i) **Climate and Natural Forces:** The result of natural forces like climate is causing seasonal variation. Umbrellas are sold more in rainy season. In winter season, sale of the woolen clothes will increase. In hot season, the sales of ice, ice-cream, fruit salad, etc., will increase. Thus climate and weather play an important role in seasonal movement. Agricultural production depends upon the monsoon.

ii) **Customs and Habits:** Man-made conventions are the customs, habits, fashion, etc. There is the custom of wearing new clothes, preparing sweets and buying crackers for Deepavali, Onam, Christmas, etc. At that time, there is more demand for cloth, sweets and crackers. It will happen every year. In marriage season, the price of gold will increase.

Purposes of Studying Seasonal Variations

i) To analyze seasonal pattern in a short-period time series.

ii) Once the seasonal factor is known it can be used for separating cyclical and irregular forces from the residual forces of changes.

iii) The seasonal factor can be used for adjustment in the value projected on the basis of trend. Short-term forecasts are always affected by seasonal factors.

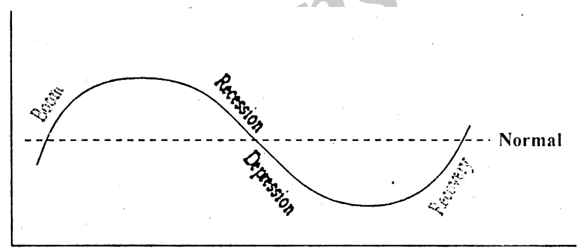
iv) The observed value can be appraised in terms of seasonal factor and adjusted so as to get correct idea of the trend, if any, in an economic phenomenon. The study of seasonal variations is, therefore, necessary for proper appraisal of business facts influenced by a seasonal variation.

v) The seasonal factor can also be used for pricing of articles and services so as to level up the seasonal variations in demand.

vi) A study of the seasonal patterns is extremely useful in planning future operations and in formulation of policy decisions regarding purchase, production, inventory control, personnel requirement, selling and advertising programme, etc.

3) Cyclical Variations (C)

As the economy expands during a period of boom, we would expect to find that such data as sales, output or consumer expenditure also show a rising trend; and during a period of slump, we would expect them to show a downward trend. Thus a wavelike motion may be observed in the pattern of our data. They are prosperity (boom) recession, depression and recovery; and are shown below:



Business Cycle

The study of cyclical variation is helpful to businessmen and economist for framing suitable policies and stabilizing the level of business activities.

Purposes of Studying Cyclical Variation

i) One can easily study the character of business fluctuations. Good policies can be formulated at stabilizing the level of business activity.

ii) Businessman can take timely steps in maintaining business during booms and depressions.

iii) A careful study of cyclical variations facilitates a businessman to face the recession period and make them ready to reap the benefits during booms.

4) Irregular Variations (I)

Irregular variations are the effect of random factors. These are generally mixed up with seasonal and cyclical variations and are caused by irregular and accidental factors like floods, famines, wars, strikes, lockouts, etc. There is no regular period or time of their occurrence and that is why they are called random or chance fluctuations. Sometimes these factors are very effective and may even give rise to cyclical fluctuations. However since these are not regular factors, no advance preparation can be done to meet their consequences. Their effects are also unpredictable and irregular.

5.2 Models of Time Series Analysis

5.2.1 Additive, Multiplicative and Mixed Models

Q4. Explain the various models are used in Time Series Analysis.

Ans :

There are two mathematical models which are commonly used for the decomposition of a time series into its components viz.

(a) Decomposition by Additive model

(b) Decomposition by Multiplicative model.

(c) Decomposition by Mixed Model

(a) Decomposition by Additive Model : According to the additive mode, a time series can be expressed as

$$Y_t = T_t + S_t + C_t + R_t$$

where Y_i = is the time series value at time t,

T_t = represent the trend value.

S_t = represents the seasonal variations.

C_t = represents the cyclic movements and

R_t = represents the random fluctuations at time *t*.

Obviously, the term S_t will not appear in a series of annual data. The additive model implicitly implies that seasonal forces (in different years), cyclical forces (in different cycles) and irregular forces (in different long time periods) operate with equal absolute effect irrespective of the long trend value. As such C_t (and S_t) will haves positive or negative values according as whether we are in an above normal or below normal phase of the cycle (and year) and the total of positive and negative values for any cycle (and any year) will be zero. R_t will also have positive or negative values and in the long run R_t will be zero. Occasionally, there may be a few isolated occurrences of extreme R_t of episodic nature.

(b) Decomposition by Multiplicative Model: On the other hand if we have reasons to assume that the various components in a time series operate proportionately to the general level of the series, the traditional or classical multiplicative model is appropriate. According to the multiplicative model

$$Y_t = T_t \times S_t \times C_t \times R_t$$

where S_t, C_t and R_t instead of assuming positive and negative values are indices fluctuating above or below unity and the geometric means of S_t in a year, C_t in a cycle and R_i in a long-term period are unity. In a time series with both positive and negative values the multiplicative model can not be applied unless the time series is translated by adding a suitable positive value. It may be pointed out that the multiplicative decomposition of a time series is same as the additive decomposition of the logarithmic values of the original time series.

In practice, most of the series relating to economic data confirm to the multiplicative model.

(c) Decomposition by Mixed Models : In addition to the additive and multiplicative models discussed above, the components in a time series may be combined in large number of ways. The different models, defined under different assumptions, will yield different results. Some of the defined models can be:

$$Y_t = (T_t \times S_t \times C_t) + R_t$$

$$Y_t = (T_t \times C_t) + (S_t \times R_t)$$

$$Y_t = T_t + (S_t \times C_t \times R_t)$$

$$Y_t = T_t + S_t + (C_t \times R_t)$$

$$Y_t = T_t \times (S_t + C_t) \times R_t$$

5.3 Trend Analysis

Q5. Define Trend Analysis. Explain the purpose of measuring trend.

Ans : **(Imp.)**

The term trend analysis refers to the concept of collecting information and attempting to spot a pattern, or trend, in the information. In some fields of study, the term "trend analysis" has more formally-defined meanings.

Although trend analysis is often used to predict future events, it could be used to estimate uncertain events in the past, such as how many ancient kings probably ruled between two dates, based on data such as the average years which other known kings reigned.

Trend analysis uses a technique called least squares to fit a trend line to a set of time series data and then project the line into the future for a forecast.

Purposes of Measuring Trend

There are three basic purposes of measuring secular trend :

1) The first purpose is to study the past growth or decline of a series. The secular trend describes the basic growth tendency ignoring short-term fluctuations.

2) The second and most important purpose of measuring secular trend is to project the curve into the future as a long-term forecast. If the past growth has been steady and if the conditions that determine this growth may reasonably be expected to persist in the future, a trend curve may be projected over five to ten years into the future as a preliminary forecast.

3) The third purpose of measuring secular trend is to eliminate it, in order to clarify the cycles and other short-term movements in the data. A steep trend may observe minor cycles. Dividing the data, by the trend values yield ratios which make the curve fluctuate around a horizontal line, thus bringing the cycles into clear relief.

Q6. Explain the uses of trend analysis.

Ans :

Uses of Trend

1) The trend describes the basic growth tendency ignoring short term fluctuations.

2) It describes the pattern of behavior which has characterized the series in the past.

3) Future behavior can be forecasted in the assumption that past behavior will continue in the future also.

4) Trend analysis facilitate us to compare two or more time series over different period of time and this helps to draw conclusions about them,.

Q7. Explain the various methods for measuring trend.

Ans :

Methods / Measurement of Trend

There are Your methods for determining trend in time series :

1. Freehand (or Graphical) Method,
2. Semi-Average Method,
3. Moving Average Method.
4. Least Squares Method;

5.3.1 Free Hand Curve

Q8. What is Free Hand Curve? Explain with illustration.

Ans :

A trend is determined by just inspecting the plotted points on a graph sheet. Observe the up and down movements of the-points. Smooth out the irregularities by drawing a freehand curve or line through the scatter points. The curve so drawn would give a general notion of the direction of the change. Such a freehand smoothed curve eliminates the short- time swings and shows the long period general tendency of the changes in the data.

Drawing a smooth freehand curve requires a personal skill and judgement. The drawn curve should pass through the plotted points in such a manner that the variations in one direction are approximately equal to the variation in other direction. Different persons, however, drawn different curves at different directions, with different slopes and in different styles. This may lead to different conclusions. To overcome these limitations, we can use the semi-average method of measuring the trend.

Example

Fit a trend line to the following data by the freehand method,

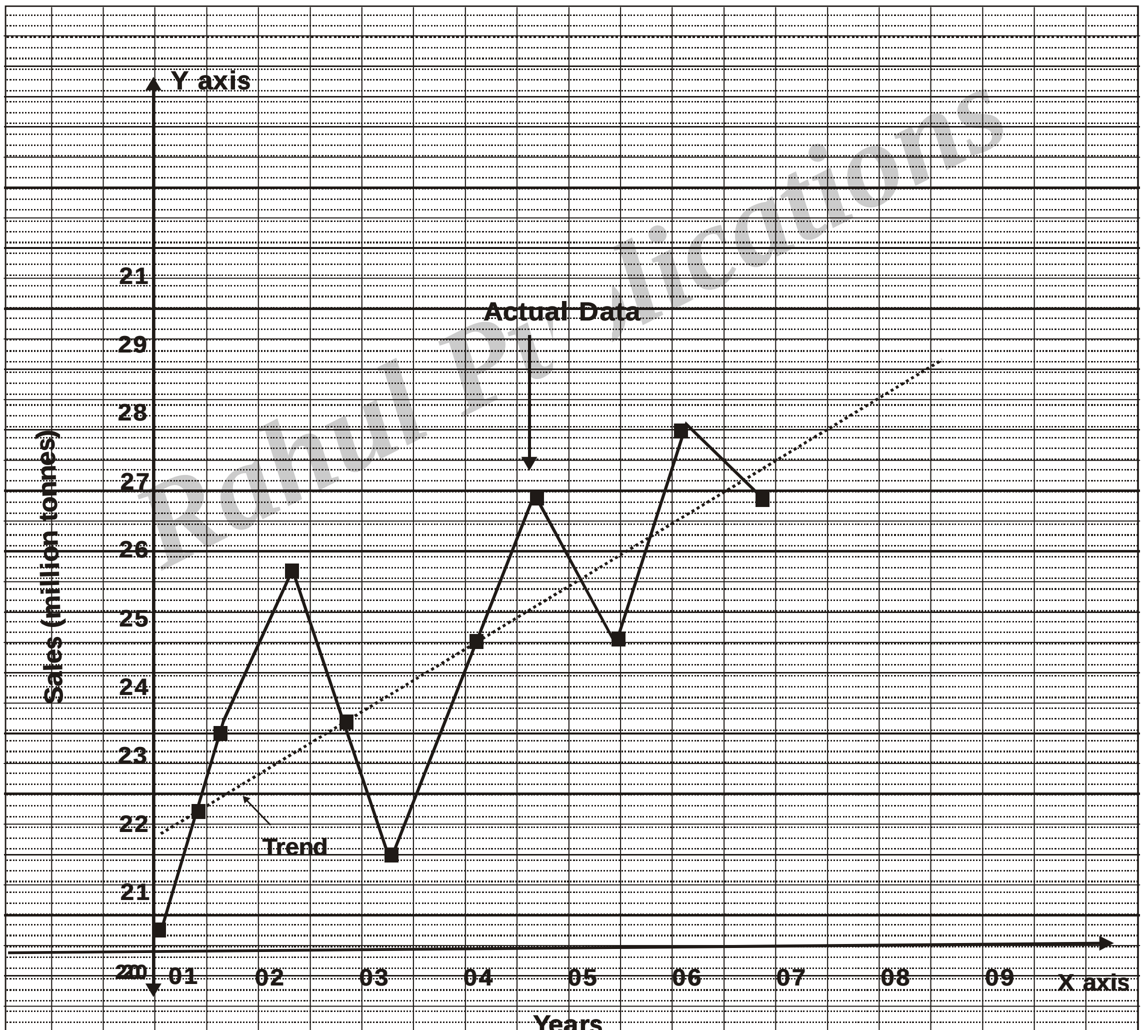
Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Sales (milllion tonnes)	19	22	24	20	23	25	23	26	25

Sol :

Steps

1. Time series data is plotted on the graph
2. The direction of the trend is examined on the basis of the plotted data (dots)
3. A straight line is drawn which shows the direction of the trend.

The actual data and the trend line are shown in the following graph.



Trend by the freehand method

5.3.2 Semi Averages

Q9. Explain the procedure of "semi-average method". What are its merits and demerits? Explain with an illustration.

Ans :

Procedure of Semi-Average Method

The following procedure is followed for semi-average method,

(i) The entire time series is classified into two equal parts with respect to time. For even period, equal split. For odd period, equal parts obtained by omitting middle period.

(ii) Compute the arithmetic mean of time series values for each half separately. These means are called semi-averages.

(iii) Semi averages are plotted as points against the middle point of the respective time period covered by each part.

(iv) The line joining these points gives the straight line trend fitting the given data.

Merits of Semi-Average Method

The following are some of the merits of semi-average method,

1. Objectivity
2. Ease of apply and understandability
3. Extend both ways the line i.e., we can get past and future estimates.

Demerits of Semi-Average Method

Some of the demerits of semi-average method are,

1. Linear trend assumption may not exist.
2. A men may be questioned.
3. Thus, values of trend are not precise and reliable.

Example

Using the following data, fit a trend line by using the method of semi-averages,

Year	1996	1997	1998	1999	2000	2001	2002
Output	700	900	1100	900	1300	1000	1600

Sol :

Step 1

The data provided in the problem is of seven years i.e., (an odd number). Thus, the middle year [1999] shall be ignored and the remaining years are divided into two equal time periods and their arithmetic averages is computed as follows,

$$\text{Average of the first three years} = \frac{700+900+1100}{3} = \frac{2700}{3} = 900$$

$$\text{Average of the last three years} = \frac{1300+1000+1600}{3} = \frac{3900}{3} = 1300$$

Therefore, the semi-averages are 900 and 1300

Step 2

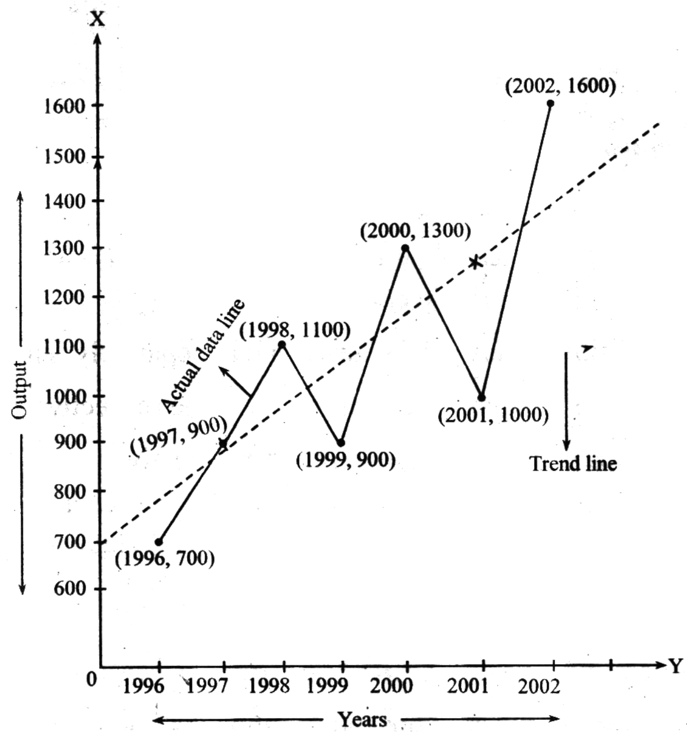
The next step is to plot the semi-averages against the mid-point (middle year) of each time period. Thus, it would , be year 1997 and 2001 respectively.

Step 3

The plotted points are joined in order to derive the trend line using the semi average method.

Step 4

The original data and the trend line is plotted on a graph as follows,



5.3.3 Moving Averages

Q10. Discuss the method of moving averages in measuring trend. What are it merits and limitations of moving average method?

Ans :

Moving Average Method

In moving average method, the average value for a number of years (month or weeks) 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 effect of averaging is to give a smoother curve, lessening the influence of the fluctuations that pull the annual figures away from the general trend.

The period of moving average is decided in the light of the length of the cycle. More applicable to data with cyclical movements.

Formula for 3 yearly moving average will be,

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3}...$$

Formula for 5 yearly moving average will be,

$$\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5}....$$

Methods of Moving Average

The following two methods are followed in moving averages,

a) Odd Yearly Method

i) Calculate 3/5...yearly totals

ii) Now compute 3/5 yearly average by dividing the totals calculated in step (i) by the respective number of years, i 3/5/...

iii) Short term oscillations are calculated using the formula, $Y - Y_C$

Where, Y - Actual value and Y - Estimated value.

b) Even Yearly Method

Example : 4 years

i) Calculate 4 yearly moving totals and place at the centre of middle two years of the four years considered.

ii) Divide 4 yearly moving totals by 4 to get 4 yearly average.

iii) Take a 2 period moving average of the moving average which gives the 4 yearly moving average centered.

Merits of Moving Average

The merits of moving average are as follows,

1. Of all the mathematical methods of fitting a trend, this method is the simplest.«
2. The method is flexible so that even if a few more observations are to be added, the entire calculations are not changed.
3. If the period of the moving average happens to coincide with the period of the cycle, the cyclical fluctuations are automatically eliminated.
4. The shape of the curve in case of moving average method is determined by the data rather than the statisticians choice of mathematical function.

Limitations of Moving Average

The following are the limitations of moving averages,

1. Trend values cannot be computed for all the years. For example, in a 5 yearly moving we cannot compute trend values for the first two and the last two years.
2. It is difficult to decide the period of moving average since there is no hard and fast rule for the purpose.
3. Moving average cannot be used in forecasting as it is not represented by any mathematical function.
4. When the trend is not linear, the moving average lies either above or below the true sweep of the data.

PROBLEMS ON MOVING AVERAGE METHOD

1. Calculate the 3-yearly moving averages of the production figures given below and draw the trend:

Year	Production (in met. tonnes)	Year	Production (in met. tonnes)
1985	15	1993	63
1986	21	1994	70
1987	30	1995	74
1988	36	1996	82
1989	42	1997	90
1990	46	1998	95
1991	50	1999	102
1992	56	–	–

Sol.

CALCULATION OF 3-YEARLY MOVING AVERAGES

Year	Production Y	3-Yearly totals	3-Yearly Moving average
1985	15	—	—
1986	21	66	22.00
1987	30	87	29.00
1988	36	108	36.00
1989	42	124	41.33
1990	46	138	46.00
1991	50	152	50.67
1992	56	169	56.33
1993	63	189	63.00
1994	70	207	69.00
1995	74	226	75.33
1996	82	246	82.00
1997	90	267	89.00
1998	95	287	95.67
1999	102	—	—

2. Construct 5-yearly moving averages of the number of students studying in a college shown below :

Year	No.of students	Year	No. of students
1990	**332**	**1995**	**405**
1991	**317**	**1996**	**410**
1992	**357**	**1997**	**427**
1993	**392**	**1998**	**405**
1994	**402**	**1999**	**438**

Sol :

CALCULATION OF 5-YEARLY MOVING AVERAGES

Year	No.of Students	5-Yearly Total	5-Yearly Moving Average
1990	332	—	—
1991	317	—	—
1992	357	1800	360.0
1993	392	1873	374.6
1994	402	1966	393.2
1995	405	2036	407.2
1996	410	2049	409.8
1997	427	2085	417.0
1998	405	—	—
1999	438	—	—

3. Assume a four-yearly cycle and calculate the trend by the method of moving averages from the following data relating to the production of tea in India :

Year	Production (mn.lbs.)	Year	Production (mn.lbs.)
1987	464	1992	540
1988	515	1993	557
1989	518	1994	571
1990	467	1995	586
1991	502	1996	612

Sol :

CALCULATION OF TREND BY THE MOVING AVERAGE METHOD

Year	Production (mn.lbs)	4-yearly moving totals	4-yearly moving average	4-yearly moving average centered
1987	464	—	—	—
1988	515	1964	491.00	—
1989	518	2002	500.50	495.75
1990	467	2027	506.75	503.62
1991	502	2066	516.50	511.62
1992	540	2170	542.50	529.50
1993	557	2254	563.50	553.00
1994	571	2326	581.50	572.50
1995	586	—	—	—
1996	612	—	—	—

4. Use of four-yearly moving average method to calculate trend for the following data and comment.

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
production	133	248	267	299	321	500	350	450	399	550

Sol :

In four yearly moving averages, the moving average values are placed in the middle of 2 time periods. An alternative is made to synchronize them with the original data by taking 2 period average of the between corresponding time periods. This technique is called centering. The corresponding moving average values are called centered moving averages.

Year (1)	(4) Production (2)	(4) Yearly Moving Totals (3)	Yearly Moving Averages (4) = (3) + 4	2 Point Moving Total (5)	(4) 4.m Average Centered (6) = (5) + 2
2005	133	–	–		
2006	248	947	236.75	521	261
2007	267	1135	283.75	631	316
2008	299	1387	346.75	715	358
2009	321	1470	367.75	773	387
2010	500	1621	405.25	830	415
2011	350	1699	424.75	862	431
2012	450	1749	437.25	–	–
2013	399	–			
2014	550	–			

5. From the following data, calculate trend values using four yearly moving average.

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007
Pr oduction (in tons)	506	620	1036	673	588	696	1116	738	663

Calculation of trend by 4-yearly moving averages

Sol :

Calculation of trend by 4-yearly moving averages

Year (1)	Production (2)	4 yearly moving Totals (3)	4 yearly moving average (4) = (3) ÷ 4	2 points moving total (5)	4 yearly moving average centred (6) = (5) ÷ 4
19999	506	–			
2000	620	–			
			2835	708.75	
2001	1036			1438	719
		2917	729.25		
2002	673		–	1477.50	738.75
2003	588	2993	748.25		
				1516.50	758.25
		3073	768.25		
2004	696			1552.75	776.375
		3138	784.5		
2005	3213			1587.75	793.875
		3213	803.25		
2006	738	–			
2007	663	–			

5.3.4 Least Square methods

Q11. Define least square method. Explain merits and demerits of least square method.

Ans : **(Imp.)**

Least square method is the most widely used method and provides us with a mathematical device to obtain an objective fit to the trend of a given time series. This method is so called because a trend line computed by this method is such that the sum of the squares of the deviation between the original data and the corresponding computed trend values is minimum. This method can be used to fit either a straight line trend or a parabolic trend.

The straight line trend equation is in the form of T = a + bX

Where, Y denotes the trend value of the dependent variable

X denotes the independent variable.

a and b are constants.

The values of a and b are obtained by solving the following normal equations.

$\Sigma Y = Na + b\Sigma X$

$\Sigma XY = a\Sigma X + b\Sigma X^2$

Where, N represents the number of years in the series.

When $\Sigma X = 0$ the above normal equations are simplified to

$$a = \frac{\sum Y}{N}$$

$$b = \frac{\sum XY}{\sum X^2}$$

By substituting a and b values in straight line trend equation Y = a + bX, we get the straight line equation which can be used for estimation of future values.

Merits of Least Squares Method

The following are the merits of least squares method,

1. The method of least squares is a mathematical method of measuring trend and is free from subjectiveness.
2. This method provides the line of best fit since it is this line from where the sum of positive and negative deviations is zero and the sum of square of deviations is the least.
3. This method enables us to compute the trend values for all the given time periods in the series.
4. The trend equation can be used to estimate the values of the variable for any given time period 'f' in future and the forecasted values are quite reliable.
5. This method is the only technique which enables us to obtain the rate of growth per annum for yearly data in case of linear trend.

Demerits of Least Squares Method

Some of the demerits of least squares are as follows,

1. Fresh calculations become necessary even if a single new observation is added.
2. Calculations required in this method are quite tedious and time consuming as compared with other methods.
3. Future predictions based on this method completely ignore the cyclical, seasonal and erratic fluctuations.
4. This method cannot be used to fit growth curves, gomper t_Z curve, logistic curve etc. to which most of the business and economic time series conform.

PROBLEMS ON LEAST SQUARE METHOD

6. Given below is the data of production of a certain company in lakhs of units.

Year	1995	1996	1997	1998	1999	2000	2001
Production	15	14	18	20	17	24	27

Compute the Linear trend by the method of least squares. predict the production for the year "2004".

Sol :

The Normal Equation are,

$$\Sigma Y = Na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

By solving these equations, we get a and b. This method is called the method of least squares.

Year	Production Y	Deviation from 1998 X	X^2	XY
1995	15	– 3	9	– 45
1996	14	– 2	4	– 28
1997	18	– 1	1	– 18
1998	20	0	0	0
1999	17	1	1	17
2000	24	2	4	48
2001	27	3	9	81
Total	**135**	**0**	**28**	**55**

Since $\Sigma X = 0$, the above equations simplify to,

$$a = \frac{\Sigma Y}{N} = \frac{135}{7}$$

$$= 19.29$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{55}{28}$$

$$= 1.96$$

$\therefore$ The straight line trend, $Y_e = 19.29 + 1.96X$.

Prediction of Production for the Year 2004

For year 2004,

$X = 6$ (deviation from mid year 1998).

$\therefore \quad Y_c = 19.29 + 1.96\,(6) = 19.29 + 11.76 = 31.05$

The production of the company for the year 2004 is predicated to be 31.05 lakhs of units.

7. The sales of a company in lakhs of rupees for the years 1994-2001 are given below:

Year	1994	1995	1996	1997	1998	1999	2000	2001
Production	550	560	555	585	540	525	545	585

Compute the Linear trend by the method of least squares. predict the production for the year "2002".

Sol :

The model of least squares helps in solving this problem.

The straight line trend is $y_c = a + bX$

By solving these normal equations, we get a and b.

Normal equation are,

$$\Sigma Y = Na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + b\,\Sigma X^2$$

Since there are even number of observations, deviations are taken from both of the middle periods i.e., 1997 and 1998.

Year	Sales Y	Deviations From mid years X	X^2	XY
1994	550	–7	49	– 3850
1995	560	–5	25	–2800
1996	555	–3	9	–1665
1997	585	–1	1	–585
1998	540	1	1	540
1999	525	3	9	1575
2000	545	5	25	2725
2001	585	7	49	4095
N = 8	**$\Sigma Y = 4445$**	**$\Sigma X = 0$**	**$\Sigma x^2 = 168$**	**$\Sigma XY = 35$**

Since $\Sigma Y = 0$, the normal equation simplify to,

$$a = \frac{\Sigma Y}{N} = \frac{4445}{8} = 555.63$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{35}{168} = 0.21$$

$\therefore$ The straight line trend is $Y_e = 555.63 + 0.21X$

Estimation of Sales for the Year 2002

For year 2002, deviation $X = 9$,

$$\therefore \quad Y = 555.63 + 0.21(9)$$
$$= 555.63 + 1.89$$
$$= 557.52$$

For year 2002, the sales of the company is estimated to be ` 557.52 lakhs.

8. Below table shows the figures of production (in thousand tons) of a sugar factory.

Year	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**	**1985**
Production	**77**	**88**	**94**	**85**	**91**	**98**	**90**

Fit a straight line trend by the method of least squares and show trend values.

Sol :

Computation of Straight Line Trend

Year (Y)	**Production**	**X**	**XY**	x^2	**Trend** $Y_c = a+bx$
1979	77	-3	-231	9	83
1980	88	-2	-176	4	85
1981	94	-1	-94	1	87
1982	85	0	0	0	89
1983	91	+1	+91	1	91
1984	98	+ 2	+196	4	93
1985	90	+ 3	+ 270	9	95
	623	0	56	28	

Since

$$\Sigma X = 0;\ a = \frac{\Sigma Y}{N} = \frac{623}{7} = 89$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{56}{28} = 2$$

Σ The equation of Trend line is $Y_C = 89 + 2x$

On the basis of the above equation, we can calculate the value of Y_c for all the seven years. For example, Y_C for 1979 = $89+2(-3)$ = 89 – 6 = 83.

1980 = 89 + 2 (–2) = 89 – 4 = 85

1981 = 89 + 2 (–2) = 89 – 2 = 87

1982 = 89 + 2 (0) = 89 + 0 = 89

1983 = 89 + 2 (1) = 89 + 2 = 91

1984 = 89 + 2 (2) = 89 + 4 = 93

1985 = 89 + 2 (3) = 89 + 6 = 95

9. Fit a straight line trend for the following series. Estimate the value for 1997.

Year	1990	1991	1992	1993	1994	1995	1995
Production of steel (in tonnes)	60	72	75	60	80	85	95

Sol. :

FITTING STRAIGHT LINE TREND

Year	Production of steel (m. tonnes) Y	Deviations from 1993 X	XY	X^2
1990	60	–3	–180	9
1991	72	–2	–144	4
1992	75	–1	–75	1
1993	65	0	0	0
1994	80	+1	+80	1
1995	85	+2	+170	4
1996	95	+3	+285	9
N = 7	**ΣY = 532**	**ΣX = 0**	**ΣXY = 136**	**$\Sigma X^2 = 28$**

The equation of the straight line trend is

$$Y_C = a + bX$$

$$\Sigma X = 0, \quad a = \frac{\sum Y}{\sum N}$$

$$= \frac{532}{7} = 76$$

and $$b = \frac{\sum XY}{\sum X^2} = \frac{136}{28}$$

$$= 4.857$$

Hence, $Y_C = 76 + 4.857 X'$

For 1997, X will be $Y_{1997} = 76 + 4.857\,(4)$

$= 95.428$

Thus, the likely sales for the year 1997 is 95.428 m tonnes

10. Fit a straight line trend by the method of least squares to the following data. Assuming that the same rate of change continues, what would be the predicted earning for the year 1998 ?

Year	1989	1990	1991	1992	1993	1994	1995	1996
Earnings (Rs. lakhs)	38	40	65	72	69	60	87	95

Sol. :

FITTING OF STRAIGHT LINE TREND BY

THE METHOD OF LEAST SQUARES

Year	Earnings (Rs.lakhs) Y	Deviations from 1992.5	Deviations multiplied by 2 X	XY	X^2
1989	38	-3.5	-7	-266	49
1990	40	-2.5	-5	-200	25
1991	65	-1.5	-3	-195	9
1992	72	-0.5	1	-72	1
1993	69	+0.5	+1	+69	1
1994	60	+1.5	+3	+180	9
1995	87	+2.5	+5	+435	25
1996	95	+3.5	+7	+665	49
N = 8	**ΣY=526**		**ΣX = 0**	**ΣXY=616**	**ΣX^2 = 168**

The equation of the straight line trend is

$$a = \frac{\sum Y}{N} = \frac{526}{8} = 65.75$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{616}{168} = 3.667 \qquad \boxed{Y_c = 65.75 + 3.667\,(X)}$$

For 1998, X will be +11.

When X is 11, Y will be

$$Y_C = 65.75 + 3.667(11) = 65.75 + 40.337 = 106.087.$$

Thus, the estimated earnings for the year 1998 are Rs.106.087 lakhs.

11. The following data relate to the number of passenger cars (in millions) sold from 2004 to 2011.

Years	2004	2005	2006	2007	2008	2009	2010	2011
Number	6.7	5.3	4.3	6.1	5.6	7.9	5.8	6.1

a) Fit a straight line trend to the data through 2009.

b) Use your result in, (a) To estimate production in 2011 and compare with the actual production.

Sol :

Let, the equation of the straight line trend be $Y_C = a + b_x$, then,

$\Sigma Y = Na + b\Sigma X$

$\Sigma XY = a\Sigma X + b\Sigma X$

a) Calculations to fit a straight line

Year	No. of Passenger cars (in millions) (Y)	Deviations from 2009	X^2 (X)	XY
2004	6.7	-5	25	-33.5
2005	5.3	-4	16	-21.2
2006	4.3	-3	9	-12.9
2007	6.1	-2	4	-12.2
2008	5.6	-1	1	-5.6
2009	7.9	0	0	0
2010	5.8	1	1	5.8
2011	6.1	2	4	12.2
N = 8	**ΣY = 47.8**	**ΣX = -12**	**ΣX^2 = 60**	**ΣXY = –67.4**

$\Sigma Y = Na + b\Sigma X$ or $47.8 = 8a - 126$ ... (1)

$\Sigma XY = a\Sigma X + b\Sigma X^2$ or $-67.4 = -12a + 60b$ ... (2)

Simplifying both the equations by multiplying equation(l) by 3 and equation (2) by 2, we get.

$$143.4 = 24a - 36b$$
$$-134.8 = -24a + 120b$$
$$8.6 = 84b$$

i.e., $b = \frac{8.6}{84} = 0.102$

Substituting the value of b in equation (1), we get,

$47.8 = 8a - 12\,b$

$47.8 = 8a - 12\,(0.102)$

$47.8 = 8a - 1.224$

$8a = 49.024$

$$a = \frac{49.024}{8} = 6.128$$

Hence, the required straight line equation is $Y_C = 6.128 + 0.102\,X$

(b) Estimating the sale of passenger cars in 2011 by substituting the value of X = 2.

$Y_C = 6.128 + 0.102\,(2)$

$= 6.128 + 0.204 = 6.332$

Thus, the estimated sale for 2011 is 6.332 million passenger cars whereas actual sales are 6.1 million passenger cars. There is some difference in actual figure and estimated figure because estimates depend on few assumptions.

Note : As number of passenger cars sold is given in question. So, estimated sales and actual sales are calculated instead of estimated production are calculated instead of estimated production and actual production.

5.4 Index Numbers - Introduction

Q12. What do you understand by index numbers ?

Ans : **(Aug.-21)**

Some prominent definitions of index numbers are given below:

1. "Index numbers are devices for measuring differences in the magnitude of a group of related variables."

— Croxton & Cowden

2. "An index number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographic loca-tion or other characteristics such as income, profession, etc."

— Spiegel

3. "In its simplest form an index number is the ratio of two index numbers expressed as a per cent. An index number is a statistical measure—a measure designed to show changes in one variable or in a group of related variables over time, or with respect to geographic location, or other characteristic."

— Patterson

4. "In its simplest form, an index number is nothing more than a relative number, or a 'relative' which expresses the relationship between two figures, where one of the figures is used as a base."

— Morris Hamburg

5. "Generally speaking, index numbers measure the Size or magnitude of some object at a particular point in time as a percentage of some base or reference object in the past."

— Berenson & Levine

For a proper understanding of the term index number, the following points are worth considering :

(i) Index numbers are specialized averages :

As explained in the chapter on measures of central value, an average is a single figure representing a group of figures. However, to obtain an average the items must be comparable: for example, the average weight of men, women and children of a certain locality has no meaning at all. Furthermore, the unit of measurement must be the same for all the items. Thus an average of the weight expressed in kg., lb., etc., has no meaning. However, this is not so with index numbers. Index numbers are used for purposes of comparison in situations where two or more series are expressed in different units or the series are composed of different types of items. For example, while constructing a consumer price index the various items are divided into broad heads, namely (i) Food, (ii) Clothing, (iii) Fuel and Lighting, (iv) House Rent, and (v) Miscellaneous. These items are expressed in different units : thus, under the head 'food' wheat and rice may be quoted per quintal, ghee per kg., etc. Similarly, cloth may be measured in terms of metres. An average of all these items expressed in different units is obtained by using the technique of index numbers.

(ii) Index numbers measure the net change in a group of related variables. Since index numbers are essentially averages they describe in one single figure the increase or decrease in a group of related variables under study. The group of variables may be the prices of a specified set of commodities, the volume of production in different sectors, etc. Thus, if the consumer price index of working class for Delhi has gone up to 113 in February 2000 compared to February 1999 it means that there is a net increase of 13 per cent in the prices of commodities included in the index.

(iii) Index numbers measure the effect of changes over a period of time. Index numbers are most widely used for measuring changes over a period of time. Thus we can find out the net change in agricultural prices from the beginning of First Plan period to the end of the Eighth Plan period, i.e., from 1951 to 1996. Similarly, we can compare the agricultural production, industrial production, imports, exports, wages, etc., at two different times. However, it should be noted that index numbers not only measure changes over a period of time but also compare economic conditions of different locations, different industries, different cities or different countries.

5.4.1 Characteristics of index numbers

Q13. Explain the Characteristics of Index Numbers.

Ans : **(Imp.)**

Characteristics/Features of Index Numbers

The features of index numbers are as follows,

1. **Measures the Change in Percentages**

   Index numbers measure the change in price or quantity in terms of percentages say i.e., (10%, 20%, 15%, 25% and so on). Increase or decrease in value is represented by one single figure. Like 10% increase in sales from that previous year to current year, 30% decrease in profits when compared to that of the last year and so on.

2. **Specialized Averages**

   A single figure known as "Average" is used for representing the characteristics of the complete set of data. Average acts as a basis for comparing different data sets with each other, when they have common unit of measurement of observations. When data sets do not have common unit of measurement, specialized averages of index numbers are used for the comparison.

3. **The Measured Changes cannot be Observed Directly**

   Index numbers do not measure the changes directly but it studies the relative changes or variations in factors resulting to changes-like for measuring changes in Export-Imports related factors such as available raw materials, technology, competitors etc., are studied and measured.

4. **Measures Changes in Relation to Time or Place**

   Index numbers measure change by comparing the values at different time periods or at different places like standard of living at one place is being compared with standard of living of the other place. Sales or revenue of current year is compared with that of the previous years sales or revenue.

5.4.2 Uses of Index Numbers

Q14. Explain the uses of index numbers.

Ans : **(Imp.)**

Index numbers are indispensable tools of economic and business analysis. Their significance can be best appreciated by the following points:

(i) They help in framing suitable policies.

Many of the economic and business policies are guided by index numbers. For example, while deciding the increase in dearness allowance of the employees, the employers have to depend primarily upon the cost of living index. If wages and salaries are not adjusted in accordance with the cost of living, very often it leads to strikes and lock - outs which in turn cause considerable waste of resources. The index numbers provide some guideposts that one can use in making decisions.

Though index numbers are most widely used in the evaluation of business and economic conditions, there is a large number of other fields also where index numbers are useful. For example, sociologists may speak of population indices; psychologists measure intelligence quotients which are essentially index numbers comparing a person's intelligence score with that of an average for his or her age; health authorities prepare indices to display changes in the adequacy of hospital facilities and educational research organizations have devised formulae to measure changes in the effectiveness of school systems.

(ii) They reveal trends and tendencies.

Since index numbers are most widely used for measuring changes over a period of time, the time series so formed enable us to study the general trend of the phenomenon under study. For example, by examining index number of imports for India for the last 8-10 years we can say that our imports are showing an upward tendency, i.e., they are rising year after year. Similarly, by examining the index numbers of industrial production, business activity, etc., for the last few years we can conclude about the trend of production and business activity. By examining the trend of the phenomenon under study we can draw very important conclusions as to how much change is taking place due to the effect of seasonality, cyclical forces, irregular forces, etc. Thus index numbers are highly useful in studying the general business conditions.

(iii) They are important in forecasting future economic activity.

Index numbers are useful not only in studying the past and present workings of our economy, but they are also important in forecasting future economic activity. Index numbers then are often used in time series analysis, the historical study of long-term trend, seasonal variations and business cycle development, so that business leader may keep pace with changing economic and business conditions and have better information available for decision-making purposes.

(iv) Index numbers are very useful in deflating.

Index numbers are highly useful in deflating, i.e., they are used to adjust the original data for price changes, or to adjust wages for cost of living changes and thus transform nominal wages into real wages. Moreover, nominal income can be transformed into real income and nominal sales into real sales through appropriate index numbers.

Q15. Explain the problems in the construction of index number.

Ans : **(Jan.-20, Imp.)**

1. The Purpose of the Index

At the very outset the purpose of constructing the index must be very clearly decided - what the index is to measure and why? There is no all-purpose index. Every index is of limited and particular use. Thus, a price index that is intended to measure consumers' prices must not include wholesale prices. And if such an index is intended to measure the cost of living of poor families, great care should be taken not to include goods ordinarily used by middle class and upper-income groups. Failure to decide clearly the purpose of the index would lead to confusion and wastage of time with no fruitful results. Other problems such as the base year, the number of commodities to be included, the prices of the 'commodities, etc., are decided in the light of the purpose for which the index is being constructed.

2. Selection of a Base Period

Whenever index numbers are constructed a reference is made to some base period. The base period of an index number (also called the reference period) is the period against which comparisons are made. It may be a year, a month or a day. The index for base period is always taken as 100. Though the selection of the base period would primarily depend upon the object of the index, the following , points need careful consideration of base period.

3. Selection of Number of Items

The items included in an index should be determined by the purpose for which the index is constructed. Every item cannot be included while constructing an index number and hence one has to select a sample. For example, while constructing a price index it is impossible to include each and every commodity. Hence it is necessary to decide what commodities to include. The commodities should be selected in such a manner that they are representative of-the tastes, habits and customs of the people for whom the index is meant.

4. Price Quotations

After the commodities have been selected, the next problem is to obtain price quotations for these commodities. It is a well known fact that prices of many commodities vary from place to place and even from shop to shop in the same market. It is impracticable to obtain price quotations from all the places where a commodity is dealt in. A selection must be made of representative places and persons.

5. Choice of an Average

Since index numbers are specialized averages a decision has to be made as to which particular average (i.e. arithmetic mean, median, mode, geometric mean or harmonic mean) should be used for constructing the index.

6. Selection of Appropriate Weights

The problem of selecting suitable weights is quite important and at the same time quite difficult to decide. The term ‘weight’ refers to the relative importance of the different items in the construction of the index. All items are not of equal importance and hence it is necessary to devise some suitable method whereby the varying importance of the different items is taken into account. This is done by allocating weights. Thus, we have broadly two types of indices unweighted indices and weighted indices.

7. Selection of an Appropriate Formula

A large number of formulae have been devised for constructing the index. The problem very often is that of selecting the most appropriate formula. The choice of the formula would depend not only on the purpose of the index but also on the data available. Prof. Irving Fisher has suggested that an appropriate index is that which satisfies time reversal test and factor reversal test. Theoretically. Fisher’s method is considered as “ideal” for constructing index number. However, from a practical point of view there are certain limitations of this index which shall be discussed later. As such, no one particular formula can be regarded as the best under all circumstances. On the basis of this knowledge of the characteristics of different formulae, a discriminating investigator will choose technical methods adapted to his data and appropriate to his purposes. None of the above problems is simple to solve in practice and the final index is usually the product of compromise between theoretical standards and the standards attainable with the given data.

Q16. Explain the limitations of index numbers

Ans :

Index Numbers are of great use in studying trends, events, pertaining to a defined phenomenon. However, they suffer from certain limitations. There are

1. Approximate Representation

Index Numbers are based on sample data. Hence, they are only approximate indicators. They may not fully reflect the changes in the relative level of a phenomenon.

2. Likelihood of Error

There is a possibility of error in the choice of the sample, selection of base period, data collection, and assignment of weights or the appropriate formula to be used.

3. Not Responsive

There are rapid changes in Technology, tastes, fashions, customs and consequently, the consumption pattern of various commodities. Index numbers may not be able to keep pace with such changes. Hence, they may not be able to reflect the changes in the phenomenon being studied.

4. No Universally Acceptable formula

There is no universally acceptable standard formula that can capture the changes in a phenomenon with perfection. None of the formulae seems to satisfy all the tests of consistency. Hence, some amount of formula error is present in any calculation of Index numbers.

5. Scope for manipulation

On account of wide variety of formulae index numbers can be manipulated by appropriate choice of base year, sample set, price and quantity quotations.

6. Other Limitations

Apart from the above, Index numbers cannot capture the aspects of quality, reliability service etc. There is no index that is universally applicable. An Index constructed for one purpose cannot be used for another. Index numbers being averages, they are subject to all limitations of an average. Lastly, the utility of Index Numbers is limited by the availability of adequate and accurate data.

5.4.3 Types of Index Numbers

5.4.3.1 Unweighted Price Indexes Weighted Price Indexes

Q17. Explain the various methods of constructing index numbers.

Ans : **(Imp.)**

The various methods of constructing index, numbers are shown in the following figure:

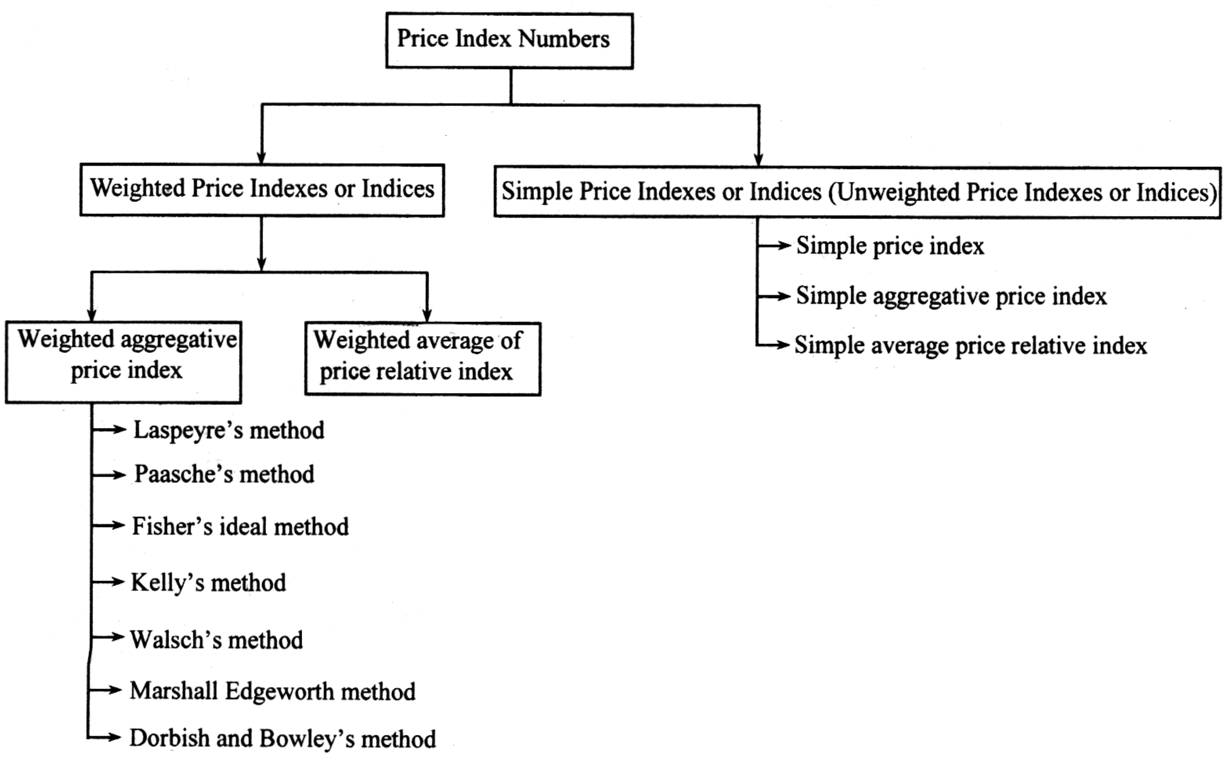


Fig : Types and Methods of Price Index Numbers

1. Weighted Price Indexes

At the time of constructing the weighted price indexes or indices, the rational weights are allocated in an explicit manner. These rational weights show the relative significance of items or commodities which are related with the computation of an index. Quantity weights and value weights are used in this weighted indexes or indices. Weighted price indexes, or indices are further divided into two types as follows,

(a) Weighted aggregate price index

(b) Weighted average of price relative index.

(a) Weighted Aggregate Price Index

In a weighted aggregate price index, certain weight is assigned to each and every commodity or item of group in accordance with its significance. This helps in gathering more information and improving accuracy of the estimates. The following methods are used in weighted aggregate price index:

(i) Laspeyre's method

(ii) Pasche's method

(iii) Fisher's ideal method

(iv) Kelly's method

(v) Walsch's method

(vi) Marshal Edgeworth's method

(vii) Dorbish and Bowley's method

(b) Weighted Average of Price Relative Index

In weighted average of price relative index, value of each commodity or item related with the calculation of composite index is ascertained by multiplying the price of each item with its quantity consumed. Quantity consumed is considered for computing the weighted average of price relative. The formula for weighted average of price relative index is as follows,

$$P_{01} = \frac{\Sigma\left(\left(p_1 \div p_0\right) \times 100\right)\left(p_0 q_0\right)}{\Sigma p_0 q_0} = \frac{\Sigma PV}{\Sigma V}$$

$$= \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

Where,

$V(= p_0 q_0) =$ Base prices and quantities determining values

$P\left(= \frac{p_1}{p_0} \times 100\right) =$ Price relative

This formula is equivalent to the formula of Laspeyre's index formula,

If 'V' is taken as p_0, q_1, then the formula would,

$$P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

Then it would be equal to Paasche's index method.

2. Unweighted Price Indexes or Indices

The unweighted indexes or indices include the following methods,

(a) Simple price index

(b) Aggregate price index

(c) Average price relative index

(a) Simple/Single Price Index

Single price index is computed by dividing the current year price of the commodity with its base year price. It is a percentage ratio which represents the comparison of a particular commodity price. The general formula used for single price index is as follows,

$$\text{Single price index in period 'n'} = \frac{P_n}{P_0} \times 100$$

Where,

P_n = Price of the commodity in the n^{th} year

P_0 = Price of the commodity in the base year

(b) Simple Aggregate Price Index

In aggregate price index, the sum of current year prices of various commodities is divided with the sum of base year prices of that various commodities. The formula is given as follows,

$$\text{Aggregate price index, } P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

Where,

P_1 = Unit price of a current year prices of all commodities

P_0 = Sum of base year prices of all commodities

(c) Simple Average Price Relative Index

The method is an improvement over the aggregate price method is,

$$P_{01} = \sum\left(\frac{P_1}{P_0} \times 100\right)$$

Where,

n = Number of commodities included in the computation of the index.

Q18. Explain Laspeyre's Method of index numbers.

Ans :

Laspeyre's Method:

This method takes the quantities of the commodities in the base period as the weight of that commodity for the purpose of calculating the index numbers. The following steps may be following.

Step I:

Multiply the current year price (represented by p_0) with the quantities of the base year (q_0) for each commodity.

Step 2 :

Add the numbers obtained in step 1. The resultant sum is represented as $\Sigma p_1 q_0$

Step 3:

Multiply the prices of base year (represented by p_0) with the quantities of the base year for each commodity.

Step 4 :

Add the numbers obtained in step 3. The resultant sum is represented as $\Sigma p_1 q_0$

Step 5 :

The index number as per Laspeyre's method

$$= P_{0,1} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

Q19. Explain Paasche's Method of index number.

Ans :

Paasche's Method:

Paasche's Method is similar to Laspeyre's method. The only difference is in assignment of weights. As per this method quantities consumed of the commodities in the current year is taken as basis. The following steps need to be followed.

Step 1 :

Multiply current year's prices (p_1) with current year's quantities (q_1)

Step 2 :

Add the numbers obtained in step (1). The resultant sum is $\Sigma p_1 q_1$

Step 3 :

Multiply base year's Prices (p_0) with current year's quantities (q_1)

Step 4 :

Add the numbers obtained in step (2). The resultant sum is $\Sigma p_0 q_1$

Step 5 :

Index Number as per Paasche's method $= P_{0,1} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

Q20. Explain Fisher's Ideal Index

Ans :

Fisher's Ideal Index :

This is the most popular amongst all weighted aggregative index numbers. It is obtained by calculating the Geometric Mean (G.M) of Laspeyre's and Paasche's index numbers. The formula for calculating fisher's ideal index is an under.

$$P_{0.1} = \left[\sqrt{\frac{\Sigma p_1 q_0 \times \Sigma p_1 q_1}{\Sigma p_0 q_0 \times \Sigma p_0 q_1}}\right] \times 100$$

Reasons for Fisher's Index being called an Ideal Index:

We noticed that Fisher's index is called an Ideal index. The reasons for the same are as follows:

1. It gives weightage to both current consumption and base year consumption.
2. It is free from upward or downward bias.
3. It satisfies both time reversal and factor reversal tests (to be discussed later)
4. It is a Geometric mean of Laspeyre's index and Paasche's index.

Q21. Define :

(a) Kellye's price index

(b) Walsh price index

(c) Dorbish Bowley price index

Ans :

(a) Kellye's price index

Kelly's Method emphasizes that the weights chosen for construction of an index number should be constant. While quantities are taken as weight, such quantities need not pertain to either base period or current period. The quantities may be an arithmetic or geometric mean of quantities of two or three years. They may also belong to some landmark year in recent history. The advantage of this method is that the weights need not be changed even if there is a change in base year. The formula for calculating Kelly's index is as follows.

$$P_{0.1} = \frac{\Sigma p_1 q}{\Sigma p_0 q} \times 100$$

Where q = fixed quantity weights

(b) Walsh price index

This method takes the Geometric Mean of quantities of base year and current year as weights. It is not a very popular method. The formula for calculating Walsh Index is an under.

$$P_{0.1} = \frac{\Sigma p_1 w}{\Sigma p_0 w} \times 100, \text{ Where } q = \sqrt{q_0 q_1}$$

(c) Dorbish Bowley price index

This method is a combination of Laspeyre's method and paasche's method. It is obtained by Calculating the Arithmetic mean of laspeyre's and Paasche's index numbers. It is also called 'L-P' formula. Thus, the index number is

$$P_{0.1} = \frac{1}{2}\left[\left(\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}\right) + \left(\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}\right)\right] \times 100$$

PROBLEMS ON INDEX NUMBER

12. From the following data, construct an Index number for 1994 taking 1999 as base as per simple Aggregative method

Commodities	Price in 1993 (Rs)	Price in 1994 (Rs)
A	40	60
B	60	90
C	85	125
D	25	35
E	30	40

Sol :

Construction of Price Index

Commodities	Price in 1993 in Rs. (P_o)	Price in 1994 in Rs. (P_1)
A	40	60
B	60	90
C	85	125
D	25	35
E	30	40
	$\Sigma P_o = 240$	$\Sigma P_1 = 350$

$$P_{0,1} = \left(\frac{\Sigma p_1}{\Sigma p_0}\right) \times 100 = \frac{350}{240} \times 100 = 145.83$$

13. Following are the prices of commodities in 1970 and 1975. Calculate a price index based on price relatives using the arithmetic mean as well as geometric mean.

Year	Commodity					
	A	B	C	D	E	F
1970	45	60	20	50	85	120
1975	55	70	30	75	90	130

Sol :

(a) Arithmetic Mean :

Commodity	P_0 (Price in 1970)	P_1 (Price in1971)	Price Relative P	Log P
A	45	55	55/45 × 100 = 122.22	2.0871
B	60	70	116.67	2.0668
C	20	30	150.00	2.1761
D	50	75	150.00	2.1761
E	85	90	105.88	2.0245
F	120	130	108.33	2.0346
			ΣP = 753.10	**12.565.2**
				= ΣLogp

$$\text{Average of Price Relatives} = \frac{753.10}{6} = 125.52$$

(b) Geometric Mean

$$P_{0,1} = \text{Antilog}\left[\frac{\Sigma \text{Logp}}{N}\right] = \text{Antilog}\left(\frac{12.5652}{6}\right) = \text{Antilog } (2.0942) = 124.3$$

14. Construct index numbers of price from the following data by applying :

1. **Laspeyres method**
2. **Paasche method**
3. **Fisher's ideal method**

	1999		2000	
Commodity	**Price**	**Quantity**	**Price**	**Quantity**
A	**2**	**8**	**4**	**6**
B	**5**	**10**	**6**	**5**
C	**4**	**14**	**5**	**10**
D	**2**	**19**	**2**	**13**

Sol :

Calculation of Various Indices

Commodity	1999		2000					
	Price p_0	Qty q_0	Price p_1	Qty q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	2	8	4	6	32	16	24	12
B	5	10	6	5	60	50	30	25
C	4	14	5	10	70	56	50	40
D	2	19	2	13	38	38	26	26
					200	160	130	103

1. **Laspeyres Method :** $P_{0,1} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$; where $\Sigma p_1 q_0 = 200$, $\Sigma p_0 q_0 = 160$

$$P_{0,1} = \frac{200}{160} \times 100 = 125$$

2. **Paasche's Method :** $P_{0,1} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$; where $\Sigma p_1 q_1 = 130$, $\Sigma p_0 q_1 = 103$

$$P_{0,1} = \frac{130}{103} \times 100 = 126.21$$

3. **Fisher's Ideal Method:** $P_{0,1} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100 = \sqrt{\frac{200}{160} \times \frac{130}{103}} \times 100$

$= \sqrt{1.578} \times 100 = 1.256 \times 100 = 125.6$

15. From the following data construct a price index number of the group of four commodities by using the appropriate formula :

Commodity	Base Year		Current Year	
	Price per unit	Expenditure (Rs.)	Price per unit	Expenditure (Rs.)
A	2	40	5	75
B	4	16	8	40
C	1	10	2	24
D	5	25	10	60

Sol :

Since we are given the base year and current year prices and expenditure, Fisher's ideal. Formula shall be most appropriate

Construction of Price Index

Commodity	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	2	20	5	15	100	40	75	30
B	4	4	8	5	32	16	40	20
C	1	10	2	12	20	10	24	12
D	5	5	10	6	50	25	60	30
					202	91	199	92

$$P_{0,1} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92}} \times 100 = 2.1912 \times 100 = 219.12$$

16. Compute Index Numbers from the following data for 2010 with 2005 as base by (i) Lasperyes Method (ii) Pasche's Method and (iii) Fisher's Ideal formula.

Commodity	2005		2010	
	Price	Qty	Price	Qty
A	20	8	40	6
B	50	10	60	5
C	40	15	50	15
D	20	20	20	25

Sol :

Comm's	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	20	8	40	6	320	160	240	120
B	50	10	60	5	600	500	300	250
C	40	15	50	15	750	600	750	600
D	20	20	20	25	400	400	500	500
					2070	**1600**	**1790**	**1470**

i) Laspayer's Method

$$P_{0,1} = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100 = \frac{2070}{1660} \times 100 = 124.69$$

ii) Paasches Method

$$P_{0,1} = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times 100 = \frac{1790}{1470} \times 100 = 121.77$$

iii) Fisher's ideal formula

$$P_{0,1} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$$

$$= \sqrt{\frac{2070}{1660} \times \frac{1790}{1470}} \times 100$$

$$= \sqrt{1.2469 \times 1.2176} \times 100$$

$$= \sqrt{1.5182} \times 100$$

$$= 1.2321 \times 100$$

$$= 123.21.$$

17. From the following data, calculate Price Index numbers by (i) Laspeyre's (ii) Paasche's and (iii) Fisher's ideal method.

Commodity	2005		2009	
	Price `	Qty.	Price `	Qty.
A	10	4	20	3
B	25	5	30	3
C	20	8	25	8
D	10	10	10	13

Sol :

Calculate price index

Commodity	2005		2009		P_1q_0	$P_0\,q_0$	P_1q_1	$P_0\,q_1$
	p_0	q_0	p_1	q_1				
A	10	4	20	3	80	40	60	30
B	25	5	30	3	150	125	90	75
C	20	8	25	8	200	160	200	450
D	10	10	10	13	100	100	130	100
					530	425	480	655

i) Laspeyres ($P_{0,1}$) $= \dfrac{\sum P_1q_0}{\sum P_0q_0} \times 100$

$= \dfrac{530}{425} \times 100$

$= 124.70$

ii) Paasches ($P_{0,1}$) $= \dfrac{\sum P_1q_1}{\sum P_0q_1} \times 100$

$= \dfrac{480}{655} \times 100$

$= 73.28$

iii) Fishers ideal ($P_{0,1}$) $= \sqrt{\dfrac{\sum P_1q_0}{\sum P_0q_0} \times \dfrac{\sum p_1q_1}{\sum p_0q_1}} \times 100$

$= \sqrt{\dfrac{530}{425} \times \dfrac{480}{655}} \times 100$

$= \sqrt{1.247 \times 0.732} \times 100$

$= 0.9554 \times 100$

$= 95.54$

5.4.4 Tests of Adequacy

Q22. Explain the criterion for testing the consistency of good index numbers.

Ans :

1. Test

This test states that the formula of index number should be independent of the units in which the prices or quantities of various commodities (or items) are quoted. All the formulae, except the index number based on simple aggregate of prices (quantities) satisfy this test.

2. Time Reversal Test

This test was proposed by Prof. Irwin Fisher. According to fisher 'the formula for calculating the 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 the base, or putting it another way, the index number reckoned forward should be reciprocal of the one reckoned backward.

In simple terms, given two time periods I and II, if an index number is calculated for period II taking period I as base, its value should be the reciprocal value of the index number for period I taking period II as base. The index numbers for the purpose of the test, should be in decimal form and not in percentage form. In other words, $P_{0,1}$ and $P_{1,0}$ should not be multiplied with 100. Symbolically.

$$\boxed{P_{0,1} \times P_{1,0} = 1}$$

Where $P_{0,1}$ = Fishers index for period II taking period I as base and

$P_{0,1}$ = Fishers index for period I taking Period II as base

Time Reversal test is satisfied by Marshall-Edge worth, Fisher, Walsh. Kelly' s index numbers and also by simple Aggregative Index, simple geometric mean of price relatives and weighted average of price relatives. Laspeyre's and Paasche's index numbers do not satisfy the time reversal test.

3. Factor Reversal Test : This test was also proposed by proof Fisher. This test requires that the product of two index numbers, one measuring price taking quantities as base, and the other measuring quantities taking price as base, should be equal to the net increase in total value from one period to another. Let us illustrate the same with the help of an example.

If $P_{0,1}$is price index number, $Q_{0,1}$ is quantity index number, the product of the two index numbers should be equal to the value index number $V_{0,1}$

$$\boxed{P_{0,1} \times Q_{0,1} = V_{0,1}}$$

Fishers index number satisfies the factor reversal test. No other method satisfies the factor reversal test.

4. Circular Test: The circular test was proposed by Weztergaard. It is an extension of the time-reversal test. It more than two time periods are considered, price index is calculated for each period with the previous year as base period. Lastly, the price index for the first year is calculated taking the last period as the base. The product of all the price index numbers should be equal to 1. Symbolically, if three time peril are considered,

$$\boxed{P_{0,1} \times P_{1,2} \times P_{2,1} = 1}$$

Only simple geometric mean of price relatives method and Kelly's method Satisfy the circular test.

PROBLEMS

18. Compute Fisher's Index Number and apply both factor reversal test and time reversal test from the following data :

Commodity	Base year		Current year	
	Price Rs.	Expenditure Rs.	Price Rs.	Expenditure Rs.
A	5	25	10	60
B	1	10	2	24
C	4	16	8	40
D	2	40	5	75

Sol :

Comm's	P_0	q_0	P_1	q_1	P_1q_0	P_0q_0	P_1q_1	P_0q_1
A	5	5	10	6	50	25	60	30
B	1	10	2	12	20	10	24	12
C	4	4	8	5	32	16	40	20
D	2	20	5	15	100	40	75	30
					202	91	199	92

Fishers Ideal Index :

$$P_{0,1} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}} \times 100$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92}} \times 100$$

$$= \sqrt{\frac{40,198}{8372}} \times 100 = \sqrt{4.8015} \times 100$$

$$= 2.191 \ \text{x} \ 100 = 219.1$$

Time Reversal Test :

$$P_{0,1} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}}$$

$$P_{1,0} = \sqrt{\frac{\sum p_0q_1}{\sum p_1q_1} \times \frac{\sum p_0q_0}{\sum p_1q_0}}$$

$$P_{0,1} \ \text{x} \ P_{1,0} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1} \times \frac{\sum p_0q_1}{\sum p_1q_1} \times \frac{\sum p_0q_0}{\sum p_1q_0}}$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92} \times \frac{92}{199} \times \frac{91}{202}} = \sqrt{1} = 1$$

$$\therefore P_{0,1} \times P_{1,0} = 1$$

Factor Reversal Test

$$P_{0,1} \ \text{x} \ q_{0,1} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1} \times \frac{\sum p_0q_1}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_1q_0}}$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92} \times \frac{92}{91} \times \frac{199}{202}} = \sqrt{\frac{199}{91} \times \frac{199}{91}} = \frac{199}{91}$$

$$V_{0,1} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} = \frac{199}{91}$$

$$\therefore P_{0,1} \times Q_{0,1} = V_{0,1}$$

19. Calculate Fisher's price index from the following data and check whether Time Reversal Test is satisfied :

Commodity	Base Year		Current Year	
	Price (Rs.)	Qty. (Kg.)	Price (Rs.)	Qty (Kg.)
A	32	50	30	50
B	30	35	25	40
C	16	55	18	50

Sol :

Calculation of Fisher's Price Index

Commodity	p_o	q_o	p_1	q_1	p_oq_o	p_1q_1	p_oq_1	p_1q_o
A	32	50	30	50	1,600	1,500	1,600	1,500
B	30	35	25	40	1,050	1,000	1,200	875
C	16	55	18	50	880	900	800	990
					3,530	3,400	3,600	3,365

Fisher's Ideal Index $= P_{0,1} = \sqrt{\frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times \frac{\Sigma p_1 q_1}{\Sigma p_o q_1}} \times 100$

$$= \sqrt{\frac{3,365}{3,530} \times \frac{3,400}{3,600}} \times 100$$

$$= \sqrt{0.95 \times 0.094} \times 100$$

$$= \sqrt{0.0893} \times 100 = 0.298 \times 100$$

$$= 29.8$$

Time Reversal Test $= P_{0,1} \times P_{1,0} = 1$

$$= \sqrt{\frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times \frac{\Sigma p_1 q_1}{\Sigma p_o q_1}} \times \sqrt{\frac{\Sigma p_o q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_o q_o}{\Sigma p_1 q_o}}$$

$$= \sqrt{\frac{3,3\cancel{65}}{3,5\cancel{30}} \times \frac{\cancel{3},400}{3,\cancel{600}}} \times \sqrt{\frac{3,6\cancel{00}}{3,\cancel{4}00} \times \frac{3,5\cancel{30}}{3,3\cancel{65}}}$$

$$= \sqrt{1} = 1$$

Hence Time Reversal Test is Satisfied.

20. Calculate Fisher's ideal index from the following data and prove that it satisfies both the time reversal and factor reversal tests :

Commodity	2000		2001	
	Price	Expenditure	Price	Expenditure
A	8	80	10	120
B	10	120	12	96
C	5	40	5	50
D	4	56	3	60
E	20	100	25	150

Sol :

Calculation of Fisher's Ideal Index

Commodity	2000		2001					
	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	8	10	10	12	100	80	120	96
B	10	12	12	8	144	120	96	80
C	5	8	5	10	40	40	50	50
D	4	14	3	20	42	56	60	80
E	20	5	25	6	125	100	150	120
					451	396	476	426

$$P_{0,1} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100 = \sqrt{\frac{451}{396} \times \frac{476}{426}} \times 100$$

$$\sqrt{1.2726} \times 100 = 1.128 \times 100 = 112.8$$

Time Reversal Test : Time reversal test is satisfied when $P_{0,1} \times P_{1,0} = 1$

$$P_{1,0} = \sqrt{\frac{\Sigma p_0q_1}{\Sigma p_1q_1} \times \frac{\Sigma p_0q_0}{\Sigma p_1q_0}} = \sqrt{\frac{426}{476} \times \frac{396}{451}}$$

$$P_{0,1} \times P_{1,0} = \sqrt{\frac{451}{396} \times \frac{476}{426} \times \frac{426}{476} \times \frac{396}{451}} = \sqrt{1} = 1$$

Hence, time reversal test is satisfied.

Factor Reversal Test : Factor reversal test is satisfied when :

$$P_{0,1} \times Q_{0,1} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}$$

$$Q_{0,1} = \sqrt{\frac{\Sigma q_1p_0}{\Sigma q_0p_0} \times \frac{\Sigma q_1p_1}{\Sigma q_1p_0}}$$

$$P_{0,1} \times Q_{0,1} = \sqrt{\frac{451}{396} \times \frac{476}{426} \times \frac{426}{396} \times \frac{476}{451}} = \frac{476}{396}$$

This is also the value of $\frac{\Sigma p_1q_1}{\Sigma p_0q_0}$. Hence, the above data also satisfies the Factor Reversal Test.

21. Construct a Fisher's ideal index from the following data and show that it satisfies time reversal and factor reversal tests :

Items	1995		1996	
	p_0	q_0	p_1	q_1
A	10	40	12	45
B	11	50	11	52
C	14	30	17	30
D	8	28	10	29
E	12	15	13	20

Sol :

Construction of Fisher's Ideal Index

Items	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	10	40	12	45	480	400	540	450
B	11	50	11	52	550	550	572	572
C	14	30	17	30	510	420	510	420
D	8	28	10	29	280	224	290	232
E	12	15	13	20	195	180	260	240
					2015	1774	2172	1914

Fisher's Ideal Index : $P_{0,1} = \sqrt{\dfrac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \dfrac{\Sigma p_1q_1}{\Sigma p_0q_1}}$

$$= \sqrt{\frac{2015}{1774} \times \frac{2172}{1914}} \times 100 = 1.135 \times 100 = 113.5$$

Time Reversal Test : Time reversal test is satisfied when :

$$P_{0,1} \times P_{1,0} = 1$$

$$P_{1,0} = \sqrt{\frac{\Sigma p_0q_1}{\Sigma p_1q_1} \times \frac{\Sigma p_0q_0}{\Sigma p_1q_0}} = \sqrt{\frac{1914}{2172} \times \frac{1774}{2015}}$$

$$P_{0,1} \times P_{1,0} = \sqrt{\frac{2015}{1774} \times \frac{2172}{1914} \times \frac{1914}{2172} \times \frac{1774}{2015}}$$

Hence, time reversal test is satisfied by the given data.

Factor Reversal Test : Factor reversal test is satisfied when :

$$P_{0,1} \times q_{0,1} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}$$

$$q_{0,1} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} = \sqrt{\frac{1914}{1774} \times \frac{2172}{2015}}$$

$$P_{0,1} \times q_{0,1} = \sqrt{\frac{2015}{1774} \times \frac{2172}{1914} \times \frac{1914}{1774} \times \frac{2172}{2015}} = \frac{2172}{1774}$$

$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$ is also equal to $\frac{2172}{1774}$. Hence factor reversal test is satisfied by the given data.

Q23. What are the problems Involved in Construction of Index Numbers

Ans :

Construction of index numbers involves resolving a few problems, which are described below :

1. Purpose of an Index

Every index is of limited and specific use. There is no all purpose Index available. Thus, great care needs to be taken to select or build an appropriate index to measure a defined phenomenon. Failure to decide clearly the purpose of the index would lead to confusion and wastage of time with no fruitful results.

2. Selection of Base Period

The base period refers to that time period with which comparisons of relative changes are made. The period that is selected as base should be a 'normal' year. In other words, it should be free from abnormalities like wars, earthquakes, famines, booms, depressions, etc. Sometimes, it is really difficult to select a year which is normal in all respects. Secondly, the base period should not be too distant in the past as decisions cannot be taken based on the situation that was prevalent very long ago. Lastly, the user must have clarity as to whether the base would be fixed or will he be following a Chain Base Index (will be explained later).

3. Constituents of an Index

The next problem is with respect to what will constitute the Index. The first issue is of sample size. In other words, how many items should be included in the Index? The sample size should neither be too small (to avoid one or two items having a huge impact on the index) nor should it be too large (to avoid the impact getting diluted). The second issue is in deciding which items would constitute the index. Ideally, the items that are most representative of the phenomenon being studied should constitute the index. Lastly, the grade or quality of the item included should be kept constant. In other words, you cannot take the price of Sona Masoori rice in one year and Basmati rice in another year and say that "rice" is a constituent of the index.

4. Price Quotations

Prices of various items included in the index may vary from one market to another. It is not practicable to obtain price quotations from all the markets. Thus, prices should be taken from markets which are large, liquid markets, facilitating a high volume of trade in that particular item. It is preferable to identify an official source or a reliable agency to provide the prices at agreed frequency.

5. Method of Calculating the Index

Once the index is constituted, the next problem is to decide the most appropriate formula for calculating the Index. Which average should be used to calculate the index? While Median, Mode and Mean are almost never used, the Geometric Mean is ideally the best but Arithmetic Mean is most popular due to its simplicity and ease of calculation. Another decision to be taken at the time of constructing the index is the usage of weights. Should simple averages be used or should weighted averages be used? If weighted averages are to be calculated, assignment of weights will involve considerable amount of skill. Lastly, a large number of methods have been presented by eminent statisticians and it is often very difficult to select the formula that is most appropriate. The Fisher's index is considered as an Ideal index but it has its own limitations.

Thus, construction of an Index is a complex process. It must be driven by the objective for which the index is being constructed.

5.5 Consumer Price Indexes

Q24. Discuss in detail consumer price index numbers/cost of living index numbers.

Ans :

Consumer Price Index Numbers

Consumer price index numbers is also known as cost of living index number is used to measure the purchasing power of a particular class of people in relation to the changes in retail prices. In other words, it studies how price variations effect the cost of living or purchasing power of a group of people.

While constructing cost of living index number, a particular section of society is selected like [Rich, middle, poor] and a study is conducted to know how price variations effect the consumption levels of that section. Based on such information Cost of Living Index Number (CLIN) is constructed.

Steps in Construction of Consumer Price Index Numbers (CPIN)

1. Selection of Group of people

A group of people or class of people is selected to construct Cost of Living Index Numbers [CLIN]. Apart from class of people, the area (i.e., rural or urban, city or town) should be clearly specified. The group of people selected for constructing cost of living index numbers must be homogenous to a maximum extent.

2. Conducting Family Budget Enquiry

An enquiry of family budget is conducted to know how much money average family spends on the consumption of different items. These items are broadly categorised into five groups namely,

(a) Food

(b) Clothing

(c) Fuel and lighting

(d) House Rent and

(e) Miscellaneous.

Each of the above groups is further sub- categorised into small groups,

Example

The group "food" is subdivided into cereals (like wheat, rice, pulses and so on) meat, fish , milk, fruits, vegetables and so on.

3. Price Quotations

While gathering information about retail prices a proper care should be taken as retail prices varies from place to place from shop to shop. Information about retail prices should be gathered from those local markets where selected class of people are located.

Uses of CPIN

1. Consumer Price Index Numbers (CPIN) are used in the preparation of wage contracts and wage negotiations.
2. CPIN assists the government and business organisation in deciding dearness allowance [D.A] to be paid to their employees.
3. CPIN are used for deflating income and value series in National Income of the country.
4. CPIN are used to measure purchasing power of money.
5. They assist in calculating Real wage by considering the variations in money income and price level.

Q25. Explain the methods of Consumer Price Index Numbers.

Ans :

There are two methods for the compute of consumer price index numbers.

(a) Aggregate Expenditure method

(b) Family Budget Method

Cost of Living Index Numbers are weighted index numbers. The commodities constitute the index are given weights according to their importance. Normally, the weights are in the ratio of amounts spent on each item. There are two methods of constructing the cost of Living Index Numbers.

a. Aggregate Expenditure Method or Weighted Aggregative :

Thus method is similar to Laspeyre's method. The quantities consumed in the base year are taken as weights. The formula is :

$$\text{Consumer Price Index} = \text{Cost of Living Index} = \frac{\Sigma p1q0}{\Sigma p0q1} \times 100$$

Since P represents price ans q represents quantity. pq is the amount spent of given commodity. Thus $\sum pq$ represents total amount spent on all items. In other words it represents total expenditure. $\sum p_0q_0$ is total expenditure incurred in the base period. $\sum p_1q_0$ is total expenditure in the current year at base period. Thus

$$\text{Consumer Price Index} = \frac{\text{Total Expenditure Current Year at Base Year Price}}{\text{Total Expenditure in Base Year}} \times 100$$

b. Family Budget Method or Method of Weighted Relatives :

The cost of living index is obtained by taking a weighted average of price relatives. The quantities consumed in the base year are taken as weights.

The formula is :

$$\text{Cost of Living Index} = \frac{\sum PV}{\sum V} \text{ Where}$$

P = $(p_1/p_0) \times 100$ for each item and V = Values Weight = 100

$$\text{Thus, Cost of Living Index} = \sum \frac{\left(\frac{p1}{p0} \times 100\right) \times p0q0}{\sum p0q0} = \frac{\sum p1q0}{\sum p0q0} \times 100$$

Thus, the cost of living index figure is one and the same, irrespective of method of construction.

PROBLEMS ON CONSUMER PRICE INDEX

22. Calculate the index number using both the Aggregate Expenditure method and Family Budget method for the year 1973 with 1960 as base year from the following data.

Commodity	Quantity in units in 1960	Price per units in 1960 (Rs)	Price per units in 1973 (Rs)
A	100	8.00	12.00
B	25	6.00	7.50
C	10	5.00	5.25
D	20	48.00	52.00
E	25	15.00	16.50
F	30	9.00	27.00

Sol :

Calculation of Consumer Price Index

Commodity	Quantity (q_0)	Price (p_0)	p_0q_0 p_1	p_0q_0 = V	p_1q_0 × 100	$P=(P_1/P_0)$	PV
A	100	8.00	12.00	800	1200	150	120000
B	25	6.00	7.50	150	187.50	125	18750
C	10	5.00	5.25	50	52.50	105	52.50
D	20	48.00	52.00	960	1040.00	108.33	1040000
E	25	15.00	16.50	375	412.50	110	41250
F	30	9.00	27.00	270	810.00	300	81000
				2605	3702.50		370205

(1) Aggregate Expenditure Method :

$$CP_1 = \frac{\sum p1q0}{\sum p0q0} \times 100 = \frac{3702.05}{2605} \times 100 = 142.13$$

(2) Family budget Method :

$$CP_1 = \frac{\sum PV}{\sum V} = \frac{370250}{2605} = 142.13$$

23. In the construction of a certain cost of living number, the following group index numbers are found. Calculate the Cost of Living Index Number by using (i) Weighted Arithmetic Mean and (ii) Weighted Geometric Mean.

Group	Index Numbers	Weights
Food	350	5
Fuel and Lighting	200	1
Clothing	240	1
House Rent	160	1
Miscellaneous	250	2

Sol :

Computation of Consumer Price Index

Group	Index No. (I)	Weights (W)	Weighted (WI)	Log I	W.log I
Food	350	5	1750	2.5441	12.7205
Fuel Lighting	200	1	200	2.3010	2.3010
Clothing	240	1	240	2.3802	2.3802
House Rent	160	1	160	2.2041	2.2041
Miscellaneous	250	2	500	2.3979	4.7958
		10	2850		24.4016

Consumer Price Index :

(i) Using Arthimethic Mean $= \frac{\sum IW}{\sum W} = \frac{2850}{10} = 285$

(ii) Using Geometric Mean $=$ Antilog $\left[\frac{\sum W \log I}{\sum W}\right] = \text{anti}\log\left[\frac{24.4016}{10}\right]$

$=$ Antilog (2.44016) $=$ 275.55

5.6 Report Writing

5.6.1 Importance of Report Writing

Q26. Define Report, Report Writing and Research Report. State the importance of Report Writing.

Ans : **(Imp.)**

i) Report

The word 'report' is originated from the Latin word "report" which implies to 'carry back'. A report is a logical presentation of facts and information. The information generated by reports is required for reviewing and evaluating progress, for planning future course of action and for taking effective decisions. Reports acts as a tool for providing feedback to the managers related to various aspects of the organization.

ii) Report Writing

Report writing is a conscious, rational and systematic effort. It is both an art as well as science. Report-writing requires conceptual and communication skills and a scientific approach to investigation, analysis and presentation. Managers should possess right abilities and attitudes for creating effective reports.

iii) Research Report

Research report is a channel to communicate the research findings to the readers of the report. A good research report is that which does its task efficiently and effectively.

Importance of Report Writting

1. **Provide Details**

   It is the research report gives every other details of research.

2. **Source of Concise and Organized Data**

   A research report is a published document and as such clear explanation is to be given for the understanding of every other reader.

3. **Logical Presentation**

   The purpose of research cannot be served unless it is presented properly.

4. **Reflects Final Research**

   Skill and care shall be taken to write a report because it is the final work of the research.

5. **Tool of Evaluating Researcher**

   It is the research report which discloses the scholarliness of the researcher.

6. **Bibliographical Evidence**

   A research report gives scope of further research and as such it is considered as bibliographical evidence.

5.6.2 Types of Research Reports

Q27. Discuss various types of Research Reports.

Ans : **(Imp.)**

Research reports vary greatly in length and type. In each individual case, both the length and the form are largely dictated by the problems at hand. For instance, business firms prefer reports in the letter form, just one or two pages in length. Banks, insurance organisations and financial institutions are generally fond of the short balance-sheet type of tabulation for their annual reports to their customers and shareholders. Mathematicians prefer to write the results of their investigations in the form of algebraic notations. Chemists report their results in symbols and formulae. Students of literature usually write long reports presenting the critical analysis of some writer or period or the like with a liberal use of quotations from the works of the author under discussion. In the field of education and psychology, the favourite form is the report on the results of experimentation accompanied by the detailed statistical tabulations. Clinical psychologists and social pathologists frequently find it necessary to make use of the case-history form.

(A) Technical Report

In the technical report the main emphasis is on (i) the methods employed, (it) assumptions made in the course of the study, (iii) the detailed presentation of the findings including their limitations and supporting data.

A general outline of a technical report can be as follows:

1. **Summary of results**

   A brief review of the main findings just in two or three pages.

2. **Nature of the study**

   Description of the general objectives of study, formulation of the problem in operational terms, the working hypothesis, the type of analysis and data required, etc.

3. **Methods employed**

   Specific methods used in the study and their limitations. For instance, in sampling studies we should give details of sample design viz., sample size, sample selection, etc.

4. **Data**

   Discussion of data collected, their sources, characteristics and limitations. If secondary data are used, their suitability to the problem at hand be fully assessed. In case of a survey, the manner in which data were collected should be fully described.

5. **Analysis of data and presentation of findings**

The analysis of data and presentation of the findings of the study with supporting data in the form of tables and charts be fully narrated. This, in fact, happens to be the main body of the report usually extending over several chapters.

6. **Conclusions**

A detailed summary of the findings and the policy implications drawn from the results be explained.

7. **Bibliography**

Bibliography of various sources consulted be prepared and attached.

8. **Technical appendices**

Appendices be given for all technical matters relating to questionnaire, mathematical derivations, elaboration on particular technique of analysis and the like ones.

9. **Index**

Index must be prepared and be given invariably in the report at the end.

The order presented above only gives a general idea of the nature of a technical report; the order of presentation may not necessarily be the same in all the technical reports. This, in other words, means that the presentation may vary in different reports; even the different sections outlined above will not always be the same, nor will all these sections appear in any particular report.

It should, however, be remembered that even in a technical report, simple presentation and ready availability of the findings remain an important consideration and as such the liberal use of charts and diagrams is considered desirable.

(B) Popular Report

The popular report is one which gives emphasis on simplicity and attractiveness. The simplification should be sought through clear writing, minimization of technical, particularly mathematical, details and liberal use of charts and diagrams. Attractive layout along with large print, many subheadings, even an occasional cartoon now and then is another characteristic feature of the popular report. Besides, in such a report emphasis is given on practical aspects and policy implications.

We give below a general outline of a popular report.

1. **The findings and their implications**

Emphasis in the report is given on the findings of most practical interest and on the implications of these findings.

2. **Recommendations for action**

Recommendations for action on the basis of the findings of the study is made in this section of the report.

3. **Objective of the study**

A general review of how the problem arise is presented along with the specific objectives of the project under study.

4. **Methods employed**

A brief and non-technical description of the methods and techniques used, including a short review of the data on which the study is based, is given in this part of the report.

5. **Results**

This section constitutes the main body of the report wherein the results of the study are presented in clear and non-technical terms with liberal use of all sorts of illustrations such as charts, diagrams and the like ones.

6. **Technical appendices**

More detailed information on methods used, forms, etc. is presented in the form of appendices. But the appendices are often not detailed if the report is entirely meant for general public.

5.6.3 Report Preparation and Presentation

Q28. Explain the process of report preparation and presentation.

Ans : **(Imp.)**

The process of preparation and presentation of report is as follows,

1. Title page
2. Table of contents
3. Foreword (Introduction)
4. Statement of objectives
5. Methodology
   (a) Research design
   (b) Data collection method
   (c) Sampling
   (d) Analysis and interpretation.
6. Limitations
7. Findings
8. Conclusions and recommendations
9. Appendix
   (a) Copies of forms used
   (b) Details of sample with validation
   (c) Tables not included in findings
   (d) Bibliography, if pertinent.

1. Title Page

The title page should indicate the subject, date on which the report is prepared for whom it is prepared and by whom. Some times it is not necessary to specify for whom the report is prepared, while at some other time, it is wise to indicate this precisely and show who actually receives copies.

Some research reports are confidential and for a limited distribution, in such cases, it is particularly desirable to indicate on the title page, about the receiver of the copy.

2. Table of Contents

If the report is lengthy or if it is divided into numerous parts, it is usually desirable to have a table of contents.

3. Foreword (Introduction)

This section serves to introduce the reader to the research project. It should give the background of the problem, the importance of the problem, the various dimensions of the problem and whether any previous research was done which is pertinent to the specific project being reported.

4. Statement of Objectives

The specific objectives of the report need to be set forth clearly. The reader must know exactly what the report covers. If the particular project is part of a large problem, it is desirable to state the overall problem and the problem solution process, sometimes it may even be wise to provide some background information as to how the problem arose and what previous research work, if any has been carried out.

If such information will help in understanding the report, it should be furnished, but it should be kept as brief as possible.

5. Methodology

The purpose of methodology section is to describe the research procedure. This includes the overall research design, the sampling procedures, the data collection method, the field methods and analysis procedure.

This section is difficult to write because it is hard to discuss methodology without using technical terms, yet much of the audience for the report will not understand technical language.

(a) Research Design : A description of the research design should make it clear, whether the study is exploratory or conclusivein nature and whether it is case, statistical, or experimental in design.

In addition, the researcher must explain why the particular design was used what its merits are for the project at hand.

(b) Data Collection Method : Whether data is collected from secondary sources or primary sources, whether results are collected by survey or observation. Again the researcher must explain why the method selected was appropriate for the project. A copy of the questionnaire or form may be included here, if the form is lengthy it can be given in the appendix, where it will not break the continuity of the report.

(c) Sampling : In describing the sampling procedure, it is first necessary to indicate the nature of universe studied. The exact sampling units, such as stores, consumers or business executives, must be defined and the geographical limits specified. If there were any difficulties in identifying the sampling units in the field, the procedure adopted for overcoming such difficulties must be explained.

(d) Analysis and Interpretation : Relatively, little can be said about the analysis and interpretation methods. The findings tend to show what has been done in this regard. If any special, statistical techniques have been used, they should be mentioned. If various executives assisted in interpretation, this fact should be noted, it helps readers appraise the interpretation given.

6. Limitations

A good report "sells" the results of the study, but it should not "oversell". Every project has limitations. The competent researcher does not attempt to gross over these points but instead calls them to the attention of the readers. This helps readers form a more accurate interpretation of the results than they would otherwise do.

It has the added advantage from the researcher's stand point, of giving confidence in the results presented. If readers find limitations which the reporter does not point out, they are apt to wonder how carefully the research was done.

7. Findings

Findings are the results of the study. This section makes up the bulk of the report. It is not just an assortment of statistical tables and charts but an organized narration of the results. Summary tables and graphical method of presentation should be used liberally. Highly detailed tables should be relegated to the appendix. The specific objectives of the study should be kept in mind while presenting the findings.

8. Conclusions and Recommendations

Conclusion should be drawn with direct reference to the objectives of the study. The reader should be able to read the objectives, turn to conclusion section and find specific conclusion relative to each objective. If, as sometimes happens, the study does not obtain satisfactory data from which to draw a conclusion relative to an objective, this should be acknowledged rather than disguised.

9. Appendixes

The purpose of the appendix is to provide a place for those report items which do not fit in the research report properly because they are either too detailed or are too specific.

For example, the appendix may contain a detailed statement of the sample design, the formulas used to determine the sampling error, detailed statistical tables, and various research forms used, such as the questionnaire and the written interviewer instructions.

Nothing shouM be relegated to the appendix if its absence from the report will make it difficult for the reader to understand the results.

Q29. What are the Factors Affecting Report Presentation?

Ans :

Following factors affect the effectiveness of the presentation :

1. Audience Analysis

If the speaker has analysed the audience in a proper way before presentation, his presentation will be more effective. On the other hand, poor or improper audience analysis leads to ineffective presentation. The style- of the presentation is largely dependent upon the type and size of the audience. If audience is large, presentation should be more formal whereas informal presentation can work in small audience.

2. Communication Environment

Communication environment affects the effectiveness of the presentations. Much of the audience notices the physical things surrounding the speaker, the stage, lighting arrangement, background, etc. Proper arrangement of these things can enhance the impact of the presentation. If there is noise in the surrounding environment, it detracts the audience from listening and consequently leaves unhealthy messages.

3. Personal Appearance

Personal appearance of the speaker has great impact on the audience. Well dressed up person can deliver good presentation. Therefore, the speaker should wear neat and clean clothes and take time to check his appearance just before starting presentation.

4. Use of Visuals

Visuals can enhance the professional image of the presentation. Different research studies demonstrate that presenters using latest visual techniques are perceived as better prepared, more persuasive, more credible and more interesting than speakers who do not use visuals. But visuals work only if the technology on which they depend works well. Therefore, presenter should check the equipment in advance before presenting."

5. Opening and Closing of Presentation

The beginning and closing of a presentation are the positions of emphasis. Those presenters who can open the presentation with interesting remarks which are likely to create more interest and enthusiasm for listening the presentation. On the other hand, presenters with poor opening are likely to leave the audience bored. Similarly, the ending of the presentation has profound impact on the audience. Endings, with vivid and positive pictures are more likely to have profound impact on the audience.

6. Organisation of Presentation

Clarity in presentation is essential that comes with proper organisation of the information. Organising the information in a proper manner can make the message more understandable, keep the audience happy and boost the image of the speaker. Proper organisation of presentation enhances the effectiveness of the presentation. On the other hand, improper organisation of the presentation will not influence the audience.

7. Language and Words

The quality of presentation is affected by the language and words. To make the audience understand the message, the speaker has to talk in the language known to the audience. To enhance the impact of presentation, he should choose the catchy words that appeal to the heart and emotions of the audience. If the language spoken by presentator is different from audience's language, and words used are stereotyped, it is likely to have least impact on the audience.

8. Quality of Voice

Quality of voice of the presenter affects the effectiveness of the presentation. Voice modulation is likely to have greater impact upon the audience whereas monotonous voice will bore the audience.

9. Body Language

The effectiveness of the presentation is also affected by the body language of the speaker. A speaker having eye contact with audience is likely to impress more than a speaker reading out the hand outs. A speaker who looked more at the audience is judged as better informed, more experienced, more honest and friendliest than a speaker who delivers the speech with less eye contact. With eye contact members of audience feel that speaker is talking to them.

Similarly, confidently moving speakers are likely to have more impact than nervous speakers. To calm one's nervousness, one should be well-prepared, take several deep breaths, relax one's muscles, pause and look at the audience and use body energy in strong gestures and movement.

10. Answering Questions

The effectiveness of presentation is also affected by presenter's skill in handling questions asked at the end of presentation. A speaker who answers the audience's questions and handles hostile questions with tact is likely to influence the audience more. On the other hand, a speaker who answers rudely will leave negative impact upon the audience.

5.6.4 Report Structure and Report Formulation

Q30. Describe the structure of an effective report.

Ans : **(Imp.)**

A report is classified into sections with headings. This classification helps in collecting information and to represent in a form which is simple and easy to read and refer to. A report is always accompanied by the covering letter.

Reports are read by different persons for fulfilling various purposes.

Few persons are interested in knowing findings and/or conclusions, while few are interested in knowing recommendations, few are interested in knowing the procedure followed for data collection etc. The details above need to be included and are required to present in a logical format.

A report can be written either by an individual or a committee. The reports are formal and must follow logical order. An individual report is personal in tone whereas committee report is impersonal in tone.

A report related to an individual is written by one person. This report follows similar logical presentation as that of committee report. The procedure to be followed when the report is short,

(i) The report is written in letter form and consists of numbered and sub-titled paragraphs.

(ii) It should be addressed to that person only who has authorized/assigned the report.

(iii) This report is a internal communication because of that it is typed on the organization's letterhead.

(iv) It is dated on the day of submission or a day earlier.

(v) The salutation consists of sir or madam as the case may be and the complimentary close is yours faithfully.

(vi) The report ends with acknowledgment of the help received in making the study when the report is long,

(a) Long report consists of complex details

(b) Long report is written in the schematic form with subheadings

(c) Long report ends with submission of the details with a covering letter.

A committee report is written by a group of persons to whom the work has been assigned. It is written in the passive voice and is impersonal in style. It is presented in a schematic form with subheadings and should not be written in the form of a letter. This report never uses the word we or our and the word committee is used if required. The committee report is typed on a plain paper and are submitted with the help of a covering letter.

The covering letter is usually typed on the organization's overhead. This type of letter is addressed to the appointing authority and clearly mention the date of submission of the report. The covering letter is signed by the convener/chairperson of the committee. The letter consists of the amount of work assigned, the completion of work and the date of submission. The covering letter is closed with a courteous message, offering to undertake further study if required.

Q31. Explain the Formulation of Research Report.

Ans :

A project report is like a road map. It is an operating document. What information and how much information it contain depends upon the size of the enterprise, as well as nature of production. For example small-scale enterprises do not include technology which is used for preparing project reports of large-scale enterprises. Within small-scale enterprises too, all information may not be homogeneous for all units. Vinod Gupta has given a general set of information in his study "Formation of a project report."

According to Gupta, project formulation divides the process of project development into eight distinct and sequential stages as below:

1. General information
2. Project description
3. Market potential
4. Capital costs and sources of finance
5. Assessment of working capital requirements
6. Other financial aspects
7. Economical and social variables
8. Project implementation

The nature of formation to be collected and furnished under each of these stages has been given below.

1. General Information : The information of general nature given in the project report includes the following:

Bio-data of promoter: Name and address, qualifications, experience and other capabilities of the entrepreneur. Similar information of each partner if any.

Industry profile: A reference analysis of industry to which the project belongs, e.g., past performance; present status, its organization, its problems etc.

Constitution and organization: The constitution and organization structure of the enterprise; in case of partnership firm its registration with registrar of firms, certification from the directorate of industries /district industry centre.

Product details: Product utility, product range, product design, advantage to be offered by the product over its substitutes if any.

2. Project Description: A brief description of the project covering the following aspects should be made in the project report.

Site: Location of the unit; owned, rented or leasehold land; industrial area; no objection certificate from municipal authorities if the enterprise location falls in the residential area.

Physical Infrastructure: Availability of the following items of infrastructure should be mentioned in the project report.

(a) Raw material: Requirement of raw material, whether inland or imported, sources of raw material supply.

(b) Skilled labour: Availability of skilled labour in the area i.e., arrangements for training labourers in various skills.

(c) Utilities: These include:

- **(i) Power:** Requirement of power, load sanctioned, availability of power
- **(ii) Fuel:** Requirement of fuel items such as coal, coke, oil or gas, state of their availability and supply position.
- **(iii) Water:** The sources of water, quality and quantity available.

(d) Pollution control: The aspects like scope of dumps, sewage system, sewage treatment plant, infiltration facility etc., should be mentioned.

(e) Communication and transportation facility: The availability of communication facilities, e.g., telephone, fax, telex, internet etc., should be indicated. Requirements for transport, mode of transport, potential means of transport, approximate distance to be covered, bottlenecks etc., should be stated in the business plan.

(f) Production process: A mention should be made for process involved in production and period of conversion from raw material into finished goods.

(g) Machinery and equipment: A complete list of machines and equipments required indicating their size, type, cost and sources of their supply should be enclosed with the project report.

(h) Capacity of the plant: The installed licensed capacity of the plant along with the shifts should also be mentioned in the project report.

(i) **Technology selected:** The selection of technology, arrangements made for acquiring it should be mentioned in the business plan.

(j) **Other common facilities:** Availability of common facilities like machine shops, welding shops and electrical repair shops etc should be stated in the project report.

(k) **Research and development:** A mention should be made in the project report regarding proposed research and development activities to be undertaken in future.

3. **Market Potential :** While preparing a project report, the following aspects relating to market potential of the product of the product should be stated in the report.

(a) **Demand and supply position:** State the total expected demand for the product and present supply position, what is the gap between demand and supply and how much gap will fill up by the proposed unit.

(b) **Expected price:** Expected price of the product to be realized should also be mentioned.

(c) **Marketing strategy:** Arrangements made for selling the product should be clearly stated in the project report.

(d) **After sales service:** Depending upon the nature of the product, provisions made for after-sales should normally be stated in the project report.

4. **Capital Costs and Sources of Finance :** An estimate of the various components of capital items like land and buildings, plant and machinery, installation costs, preliminary expenses, margin of working capital should be given in the project report. The sources should indicate the owners funds together with funds raised from financial institutions and banks.

5. **Assessment of Working Capital :** The requirement for working capital and its sources of supply should clearly be mentioned. It is preferred to prepare working capital requirements in the prescribed formats designed by limits of requirement. It will reduce the objections from banker's side.

6. **Other Financial Aspects :** To adjudge the profitability of the project to be set up, a projected profit and loss account indicating likely sales revenue, cost of production, allied cost and profit should be prepared. A projected balance sheet and cash flow statement should also be prepared to indicate the financial position and requirements at various stages of the project. In addition to this, the break even analysis should also be presented. Break even point is the level of production at which the enterprise shall earn neither profit nor incur loss. Breakdown level indicates the gestation period and the likely moratorium required for repayment of the loans. Break-even point is calculated as

Break-Even Point (BEP) = F/S – V

Where F = Fixed Cost

S = Selling Price/Unit

V = Variable Cost/Unit

The break-even point indicates at what level of output the enterprise will break even.

7. **Economical and Social Variables :** Every enterprise has social responsibility. In view of the social responsibility of business, the abatement costs, i.e., the costs for controlling the environmental damage should be stated in the project. Arrangements made for treating the effluents and emissions should also be mentioned in the report. In addition the following socio-economic benefits should also be stated in the report.

(i) Employment Generation

(ii) Import Substitution

(iii) Ancillaration

(iv) Exports

(v) Local Resource Utilization

(vi) Development of the Area

8. **Project Implimentation :** Every entrepreneur should draw an implementation scheme or a time-table for

his project to the timely completion of all activities involved in setting up an enterprise. If there is delay in implementation project cost overrun. Delay in project implementation jeopardizes the financial viability of the project, on one hand, and props up the entrepreneur to drop the idea to set up an enterprise, on the other. Hence there is need to draw up an implementation schedule for the project and then to adhere to it.

PERT and CPM discussed later in this chapter can be used to get better insight into all activities related to implementation of the project.

5.6.5 Guides for Effective Documentation

Q32. State the Guides for Effective Documentation.

Ans : **(Imp.)**

Research report is a channel of communicating the research findings to the readers of the report. A good research report is one which does this task efficiently and effectively. As such it must be prepared keeping the following precautions in view:

1. While determining the length of the report (since research reports vary greatly in length), one should keep in view the fact that it should be long enough to cover the subject but short enough to maintain interest. In fact, report-writing should not be a means to learning more and more about less and less.
2. A research report should not, if this can be avoided, be dull; it should be such as to sustain reader's interest.
3. Abstract terminology and technical jargon should be avoided in a research report. The report should be able to convey the matter as simply as possible. This, in other words, means that report should be written in an objective style in simple language, avoiding expressions such as "it seems," "there may be" and the like.
4. Readers are often interested in acquiring a quick knowledge of the main findings and as such the report must provide a ready availability of the findings. For this purpose, charts, graphs and the statistical tables may be used for the various results in the main report in addition to the summary of important findings.
5. The layout of the report should be well thought out and must be appropriate and in accordance with the objective of the research problem.
6. The reports should be free from grammatical mistakes and must be prepared strictly in accordance with the techniques of composition of report-writing such as the use of quotations, footnotes, documentation, proper punctuation and use of abbreviations in footnotes and the like.
7. The report must present the logical analysis of the subject matter. It must reflect a structure wherein the different pieces of analysis relating to the research problem fit well.
8. A research report should show originality and should necessarily be an attempt to solve some intellectual problem. It must contribute to the solution of a problem and must add to the store of knowledge.
9. Towards the end, the report must also state the policy implications relating to the problem under consideration. It is usually considered desirable if the report makes a forecast of the probable future of the subject concerned and indicates the kinds of research still needs to be done in that particular field.
10. Appendices should be enlisted in respect of all the technical data in the report.
11. Bibliography of sources consulted is a must for a good report and must necessarily be given.
12. Index is also considered an essential part of a good report and as such must be prepared and appended at the end.
13. Report must be attractive in appearance, neat and clean, whether typed or printed.
14. Calculated confidence limits must be mentioned and the various constraints experienced in conducting the research study may also be stated in the report.

15. Objective of the study, the nature of the problem, the methods employed and the analysis techniques adopted must all be clearly stated in the beginning of the report in the form of introduction.

5.6.6 Research Briefings (Oral Presentation)

Q33. What do you understand by Oral Presentation?

Ans :

At times oral presentation of the results of the study is considered effective, particularly in cases where policy recommendations are indicated by project results. The merit of this approach lies in the fact that it provides an opportunity for give-and-take decisions which generally lead to a better understanding of the findings and their implications. But the main demerit of this sort of presentation is the lack of any permanent record concerning the research details and it may be just possible that the findings may fade away from people's memory even before an action is taken. In order to overcome this difficulty, a written report may be circulated before the oral presentation and referred to frequently during the discussion. Oral presentation is effective when supplemented by various visual devices. Use of slides, wall charts and blackboards is quite helpful in contributing to clarity and in reducing the boredom, if any. Distributing a board outline, with a few important tables and charts concerning the research results, makes the listeners attentive who have a ready outline on which to focus their thinking. This very often happens in academic institutions where the researcher discusses his research findings and policy implications with others either in a seminar or in a group discussion.

Thus, research results can be reported in more than one ways, but the usual practice adopted, in academic institutions particularly, is that of writing the Technical Report and then preparing several research papers to be discussed at various forums in one form or the other. But in practical field and with problems having policy implications, the technique followed is that of writing a popular report. Researches done on governmental account or on behalf of some major public or private organisations are usually presented in the form of technical reports.

5.6.7 Referencing Styles and Citation in Business Management Research

Q34. Discuss the referencing styles and citation in business management research.

Ans : **(Imp.)**

References/Referencing Styles

References basically includes the list of documents which are usually highlighted in the research work. Example of references are the list of journal unpublished manuscripts, technical reports, articles and so on.

The list of documents contained in the reference section are usually arranged in the alphabetical order i.e., starting from the last names of the first-named authors.

The standard format of a reference is that it basically starts from the left margin of the page with succeeding lines with double space and indented.

The list of books, articles, reports and other published material which are used by the writer in his study needs to be specified either at the end of the page as a foot note (if it is one) or on a separate page under the title of references (if more than one reference is used).

There are different styles or systems of references such as Vancouver, Harvard, American Psychological Association etc. In different disciplines, different citing systems are preferred. For instance Vancouver system is preferred in medical and scientific papers and research. Two commonly used referencing styles have been discussed below.

(a) Harvard System

This system is also known as author name style where name(s) of the author(s) is/are followed by year of publication. Example

1. Prahald Mishra (2015) Business Research Method, Oxford University Press.
2. Johnson P (2009), Fundamentals of Collection Development and Management, 2nd edn, ALA Editions, Chicago.

(b) Vancouver System

This system is also known as number system where a number is assigned to each reference and the number is then embedded in the text as superscript. In the bibliography, references are listed in the number order.

Example

1. Deepak Chawla, Neena Sondhi, Research Methodology concepts and cases, Vikas Publications, 2e 2016.
2. Donald R Cooper, Pamelas. S.Schindler, Business Research Methods Tata MC Graw Hill, 2013.

Citation

Citation means the reference of the published or unpublished sources, which are used in the research work. The two main objectives of citation are as follows,

(a) To prevent plagiarism and to honestly inform the readers about the true owner of the intellectual work which is being used in the research work.

(b) To enable the reader to verily and validate the research information, by independently ascertaining the original source.

Thus, citation is a concept where in, it determines all the published and unpublished sources of the research project and its preparation.

The important parts of citation are as follows,

(i) Notes/sources cited

(ii) References

(iii) Appendix.

(i) Notes/Source Cited

The notes or sources cited are usually represented in the report as end notes or foot notes or in-text citation. These appear in the body of the research work, at the end of the page.

(ii) References

References basically includes the list of documents which are usually highlighted in the research work. Example of references are the list of journal unpublished manuscripts, technical reports, articles and so on.

(iii) Appendix

An appendix or appendices is the last element of a research paper. Appendix shows the list of attachments where they must be placed in the research paper, but it must not be placed in the body of the research paper itself.

The example of attachments which are included in the appendix section are advertisement or a brochure, a copy of questionnaire used in collecting research information, a complex mathematical table or a copy of a magazine's article, journal or newspaper. There are no specific rules related with the format of the appendix list. However, the following points are generally taken into consideration with respect to appendix.

1. The appendix list must be attached after the list of bibliography.
2. A single title page, with the heading 'Appendix' must be placed before all the attachments.
3. In case of using more than one appendix the word "Appendices" is written in the table of contents and on the section title page.
4. In the table of content, merely the number of appendix title pages is included.
5. In case of multiple appendices, they must be arranged in an alphabetical order.

Example:

Appendix A, Appendix B and so on.

Exercises Problems

1. **Compute (i) Laspeyres index and (ii) Paasches index from the data given below and test whether they satisfy (i) Time Reversal Test and (ii) Factor Reversal Tests**

Commodity	p_0	q_0	p_1	q_1
A	5	10	4	12
B	8	6	7	7
C	6	3	5	4

[Ans: $P_{0,1}$ = 83.62, $P_{0,1}$ = 119.66, $Q_{0,1}$ = 120.69, $V_{0,1}$ = 100.86

$P_{0,1}$ =83.57, $P_{1,0}$ = 119.59, $Q_{0,1}$ =120.62, $V_{0,1}$ =100.86]

2. **Convert CBI to FBI (chained to 1996) :**

Year :	1996	1997	1998	1999	2000
CBI	100	120	90	100	125

[Ans: 100, 120, 108, 108, 135]

3. **From the Chain base index numbers given below, find fixed base index numbers.**

Year :	1975	1976	1977	1978	1979
Chain Base Index :	80	110	120	90	140

[Ans: 80, 88, 105.60, 95.04, 133.06]

Short Question & Answers

1. Define Time Series.

Ans :

An arrangement of statistical data in accordance with time of occurrence or in a chronological order is called a time series. The numerical data which we get at different points of time-the set of observations is known as time series.

Examples:

i) The Annual Production of Steel in India over the last 10 years;

ii) The Monthly Sales of a Chemical Industry for the last 6 months;

iii) The daily closing price of a share in the Calculate Stock Exchange;

iv) Hourly temperature recorded by the Meteorological office in a city;

v) Yearly Price or Quantity Index Numbers.

Mathematically, a Time Series is defined by the values Y_1, Y_2, —, Y_n of a variable Y at times t_1, t_2,....t_n

Here Y is a function of time t and Y, denotes the value of the variable Y at time t. The basic assumption is that changes witnessed over time in a sample group will be extrapolated to population.

i) According to Morris Hamburg, "A time series is a set of observations arranged in chronological order".

ii) According to Kenny and Keeping, "A set of data depending on the time is called time series".

iii) According to Croxton and Cowden, "A time series consists of data arranged chronologically".

2. Define trend analysis.

Ans :

The term trend analysis refers to the concept of collecting information and attempting to spot a pattern, or trend, in the information. In some fields of study, the term "trend analysis" has more formally-defined meanings.

Although trend analysis is often used to predict future events, it could be used to estimate uncertain events in the past, such as how many ancient kings probably ruled between two dates, based on data such as the average years which other known kings reigned.

Trend analysis uses a technique called least squares to fit a trend line to a set of time series data and then project the line into the future for a forecast.

Purposes of Measuring Trend

There are three basic purposes of measuring secular trend :

1) The first purpose is to study the past growth or decline of a series. The secular trend describes the basic growth tendency ignoring short-term fluctuations.

2) The second and most important purpose of measuring secular trend is to project the curve into the future as a long-term forecast. If the past growth has been steady and if the conditions that determine this growth may reasonably be expected to persist in the future, a trend curve may be projected over five to ten years into the future as a preliminary forecast.

3. What is free hand curve?

Ans :

A trend is determined by just inspecting the plotted points on a graph sheet. Observe the up and down movements of the-points. Smooth out the irregularities by drawing a freehand curve or line through the scatter points. The curve so drawn would give a general notion of the direction of the change. Such a freehand smoothed curve eliminates the short- time swings and shows the long period general tendency of the changes in the data.

Drawing a smooth freehand curve requires a personal skill and judgement. The drawn curve should pass through the plotted points in such a manner that the variations in one direction are approximately equal to the variation in other direction.

4. Merits of least squares method.

Ans :

The following are the merits of least squares method,

1. The method of least squares is a mathematical method of measuring trend and is free from subjectiveness.
2. This method provides the line of best fit since it is this line from where the sum of positive and negative deviations is zero and the sum of square of deviations is the least.
3. This method enables us to compute the trend values for all the given time periods in the series.
4. The trend equation can be used to estimate the values of the variable for any given time period 'f' in future and the forecasted values are quite reliable.
5. This method is the only technique which enables us to obtain the rate of growth per annum for yearly data in case of linear trend.

5. Index numbers.

Ans :

Some prominent definitions of index numbers are given below:

1. "Index numbers are devices for measuring differences in the magnitude of a group of related variables."

— Croxton & Cowden

2. "An index number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographic loca-tion or other characteristics such as income, profession, etc."

— Spiegel

3. "In its simplest form an index number is the ratio of two index numbers expressed as a per cent. An index number is a statistical measure—a measure designed to show changes in one variable or in a group of related variables over time, or with respect to geographic location, or other characteristic."

— Patterson

4. "In its simplest form, an index number is nothing more than a relative number, or a 'relative' which expresses the relationship between two figures, where one of the figures is used as a base."

— Morris Hamburg

5. "Generally speaking, index numbers measure the Size or magnitude of some object at a particular point in time as a percentage of some base or reference object in the past."

— Berenson & Levine

6. Uses of index numbers.

Ans :

Index numbers are indispensable tools of economic and business analysis. Their significance can be best appreciated by the following points:

(i) They help in framing suitable policies.

Many of the economic and business policies are guided by index numbers. For example, while deciding the increase in dearness allowance of the employees, the employers have to depend primarily upon the cost of living index. If wages and salaries are not adjusted in accordance with the cost of living, very often it leads to strikes and lock - outs which in turn cause considerable waste of resources. The index numbers provide some guideposts that one can use in making decisions.

Though index numbers are most widely used in the evaluation of business and economic conditions, there is a large number of other fields also where index numbers are useful. For example, sociologists may speak of population indices; psychologists measure intelligence quotients which are essentially index numbers comparing a person's intelligence score with that of an average for his or her age; health authorities prepare indices to display changes in the adequacy of hospital facilities and educational research organizations have devised formulae to measure changes in the effectiveness of school systems.

(ii) They reveal trends and tendencies.

Since index numbers are most widely used for measuring changes over a period of time, the time series so formed enable us to study the general trend of the phenomenon under study. For example, by examining index number of imports for India for the last 8-10 years we can say that our imports are showing an upward tendency, i.e., they are rising year after year. Similarly, by examining the index numbers of industrial production, business activity, etc., for the last few years we can conclude about the trend of production and business activity. By examining the trend of the phenomenon under study we can draw very important conclusions as to how much change is taking place due to the effect of seasonality, cyclical forces, irregular forces, etc. Thus index numbers are highly useful in studying the general business conditions.

7. Consumer price index numbers/cost of living index numbers.

Ans :

Consumer price index numbers is also known as cost of living index number is used to measure the purchasing power of a particular class of people in relation to the changes in retail prices. In other words, it studies how price variations effect the cost of living or purchasing power of a group of people.

While constructing cost of living index number, a particular section of society is selected like [Rich, middle, poor] and a study is conducted to know how price variations effect the consumption levels of that section. Based on such information Cost of Living Index Number (CLIN) is constructed.

8. Define Report, Report Writing and Research Report.

Ans :

i) Report

The word ‘report’ is originated from the Latin word “report” which implies to ‘carry back’. A report is a logical presentation of facts and information. The information generated by reports is required for reviewing and evaluating progress, for planning future course of action and for taking effective decisions. Reports acts as a tool for providing feedback to the managers related to various aspects of the organization.

ii) Report Writing

Report writing is a conscious, rational and systematic effort. It is both an art as well as science. Report-writing requires conceptual and communication skills and a scientific approach to investigation, analysis and presentation. Managers should possess right abilities and attitudes for creating effective reports.

iii) Research Report

Research report is a channel to communicate the research findings to the readers of the report. A good research report is that which does its task efficiently and effectively.

9. Oral Presentation.

Ans :

At times oral presentation of the results of the study is considered effective, particularly in cases where policy recommendations are indicated by project results. The merit of this approach lies in the fact that it provides an opportunity for give-and-take decisions which generally lead to a better understanding of the findings and their implications. But the main demerit of this sort of presentation is the lack of any permanent record concerning the research details and it may be just possible that the findings may fade away from people’s memory even before an action is taken. In order to overcome this difficulty, a written report may be circulated before the oral presentation and referred to frequently during the discussion. Oral presentation is effective when supplemented by various visual devices. Use of slides, wall charts and blackboards is quite helpful in contributing to clarity and in reducing the boredom, if any. Distributing a board outline, with a few important tables and charts concerning the research results, makes the listeners attentive who have a ready outline on which to focus their thinking. This very often happens in academic institutions where the researcher discusses his research findings and policy implications with others either in a seminar or in a group discussion.

10. Referencing styles.

Ans :

References basically includes the list of documents which are usually highlighted in the research work. Example of references are the list of journal unpublished manuscripts, technical reports, articles and so on.

The list of documents contained in the reference section are usually arranged in the alphabetical order i.e., starting from the last names of the first-named authors.

The standard format of a reference is that it basically starts from the left margin of the page with succeeding lines with double space and indented.

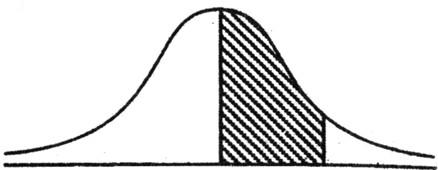
The list of books, articles, reports and other published material which are used by the writer in his study needs to be specified either at the end of the page as a foot note (if it is one) or on a separate page under the title of references (if more than one reference is used).

LOGARITHM TABLES

TABLE VI: AREAS UNDER STANDARD NORMAL PROBABILITY CURVE

Normal Probability curve is given by : $f(x) = \frac{1}{\sigma\sqrt{3\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}; -\infty < x < \infty$ and standard normal probability curve is given by :

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right), -\infty < z < \infty$$



where $Z = \frac{X - E(X)}{\sigma_x} = \frac{X-\mu}{\sigma} \sim N(0,1)$

The following table gives the shaded area in the diagram, viz., P(0 < Z < z) for different values of z.

AREAS UNDER STANDARD NORMAL PROBABILITY CURVE

z \ z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

TABLE VI A : ORDINATES OF THE NORMAL PROBABILITY CURVE

The following table gives the ordinates of the standard normal probability curve, i.e., it gives the value of

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right), -\infty < z < \infty$$

for different values of z, where:

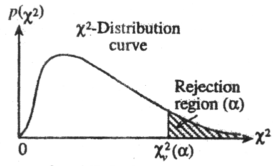
$$Z = \frac{X - E(X)}{\sigma_x} = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Obviously $\phi(-z) = \phi(z)$.

↓Z→	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.3989	.3989	.3989	.3988	.3986	.3984	.3982	.3980	.3977	.3973
0.1	.3970	.3965	.3961	.3956	.3951	.3945	.3939	.3932	.3925	.3918
0.2	.3910	.3902	.3894	.3885	.3876	.3867	.3857	.3847	.3836	.3825
0.3	.3814	.3802	.3790	.8778	.3765	.3752	.3739	.3725	.3712	.3697
0.4	.3683	.3668	.3653	.3637	.3621	.3605	.3589	.3572	.3555	.3538
0.5	.3521	.3503	.3485	.3467	.3448	.3429	.3410	.3391	.3372	.3352
0.6	.3332	.3312	.3292	.3271	.3251	.3230	.3209	.3187	.3166	.3144
0.7	.3123	.3101	.3079	.3056	.3034	.3011	.2989	.2966	.2943	.2920
0.8	.2897	.2874	.2850	.2827	.2803	.2780	.2756	.2732	.2709	.2685
0.9	.2661	.2637	.2613	.2589	.2565	.2541	.2516	.2492	.2468	.2444
1.0	.2420	.2396	.2371	.2347	.2323	.2299	.2275	.2251	.2227	.2203
1.1	.2179	.2155	.2131	.2107	.2083	.2059	.2036	.2012	.1989	.1965
1.2	.1942	.1919	.1895	.1872	.1849	.1826	.1804	.1781	.1758	.1736
1.3	.1714	.1691	.1669	.1647	.1626	.1604	.1582	.1561	.1539	.1518
1.4	.1497	.1476	.1456	.1435	.1415	.1394	.1374	.1354	.1334	.1315
1.5	.1295	.1276	.1257	.1238	.1219	.1200	.1182	.1163	.1145	.1127
1.6	.1109	.1092	.1074	.1057	.1040	.1023	.1006	.0989	.0973	.0957
1.7	.0940	.0925	.0909	.0893	.0878	.0863	.0848	.0833	.0818	.0804
1.8	.0790	.0775	.0761	.0748	.0734	.0721	.0707	.0694	.0681	.0669
1.9	.0656	.0644	.0632	.0620	.0608	.0596	.0584	.0573	.0562	.0551
2.0	.0540	.0529	.0519	.0508	.0498	.0488	.0478	.0468	.0459	.0449
2.1	.0440	.0431	.0422	.0413	.0404	.0396	.0387	.0379	.0371	.0363
2.2	.0355	.0347	.0339	.0332	.0325	.0317	.0310	.0303	.0297	.0290
2.3	.0283	.0277	.0270	.0264	.0258	.0252	.0246	.0241	.0235	.0229
2.4	.0224	.0219	.0213	.0208	.0203	.0198	.0194	.0189	.0184	.0180
2.5	.0175	.0171	.0167	.0163	.0158	.0154	.0151	.0147	.0143	.0139
2.6	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110	.0107
2.7	.0104	.0101	.0099	.0096	.0093	.0091	.0088	.0086	.0084	.0081
2.8	.0079	.0077	.0075	.0073	.0071	.0069	.0067	.0065	.0063	.0061
2.9	.0060	.0058	.0056	.0055	.0053	.0051	.0050	.0048	.0047	.0046
3.0	.0044	.0043	.0042	.0040	.0039	.0038	.0037	.0036	.0035	.0034
3.1	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026	.0025	.0025
3.2	.0024	.0023	.0022	.0022	.0021	.0020	.0020	.0019	.0018	.0018
3.3	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014	.0013	.0013
3.4	.0012	.0012	.0012	.0011	.0011	.0010	.0010	.0010	.0009	.0009
3.5	.0009	.0008	.0008	.0008	.0008	.0007	.0007	.0007	.0007	.0006
3.6	.0006	.0006	.0006	.0005	.0005	.0005	.0005	.0005	.0005	.0004
3.7	.0004	.0004	.0004	.0004	.0004	.0004	.0003	.0003	.0003	.0003
3.8	.0003	.0003	.0003	.0003	.0003	.0002	.0002	.0002	.0002	.0002
3.9	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001

TABLE VII : CRITICAL VALUES $\chi^2_\nu(\alpha)$ OF CHI-SQUARE DISTRIBUTION (RIGHT TAIL AREAS) FOR GIVEN PROBABILITY α,

Where



$$P\left[x^2 > \chi^2_\nu(\alpha)\right] = \alpha$$

AND ν IS DEGREES OF GREEDOM (d.f)

Degrees of Freedom (υ)	Probability (α)							
	.99	.975	.95	.90	.10	.05	.025	.01
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.634	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	8.231	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	23.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	23.671	35.479	38.932
22	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39..364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892

For large values of υ, the quantity $\sqrt{2\chi^2} - \sqrt{2v-1}$ may be used as a standard normal variable.

TABLE VIII : CRITICAL VALUES OF STUDENT'S t-DISTRIBUTION

d.f.	LEVEL OF SIGNIFICANCE FOR ONE-TAILED TEST						
	.25	.10	.05	.025	.01	.005	.0005
	LEVEL OF SIGNIFICNCE FOR TWO-TAILED TEST						
υ	.50	.20	.10	.05	.02	.01	.001
1	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	.816	1.886	2.920	4.303	6.965	9.925	31.599
3	.765	1.638	2.353	3.182	4.541	5.841	12.924
4	.741	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	1.476	2.015	2.571	3.365	4.032	6.869
6	.718	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	1.415	1.895	2.365	2.998	3.499	5.408
8	.706	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	1.328	1.729	2.093	2.539	2.861	3.883
20	.687	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	1.319	1.714	2.069	2.500	2.807	3.768
24	.865	1.318	1.711	2.064	2.492	2.797	3.745
25	.684	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	1.315	1.706	2.056	2.479	2.779	3.707
27	.684	1.314	1.703	2.052	2.473	2.771	3.690
28	.683	1.313	1.701	2.048	2.467	2.763	3.674
29	.683	1.311	1.699	2.045	2.462	2.756	3.659
30	.683	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	1.303	1.684	2.021	2.423	2.704	3.551
60	.679	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	.674	1.282	1.645	1.960	2.326	2.576	3.291

LOGARITHMS

											Mean Differences								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4	8	12	17	21	25	29	33	37
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19	23	26	30	34
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3	7	10	14	17	21	24	28	31
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430	3	6	10	13	16	19	23	26	29
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732	3	6	9	12	15	18	21	24	27
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14	17	20	22	25
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3	5	8	11	13	16	18	21	24
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2	5	7	10	12	15	17	20	22
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2	5	7	9	12	14	16	19	21
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989	2	4	7	9	11	13	16	18	20
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7

	0	1	2	3	4	5	6	7	8	9	Mean Differences 1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	5	5	6	7
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	6	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	4	5	6	7
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3	4	4	5	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	3	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9638	9643	9647	9653	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2	3	3	4	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	3	4

ANTI-LOGARITHMS

											Mean Differences								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
·00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021	0	0	1	1	1	1	2	2	2
·01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
·02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
·03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
·04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
·05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
·06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
·07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
·08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
·09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
·10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
·11	1288	1291	1234	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	2	2	2	2	3
·12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346	0	1	1	1	2	2	2	2	3
·13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
·14	1380	1334	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
·15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
·16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
·17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
·18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
·19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	3	3	3
·20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
·21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
·22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
·23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
·24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774	0	1	1	2	2	2	3	3	4
·25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
·26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858	0	1	1	2	2	3	3	3	4
·27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901	0	1	1	2	2	3	3	3	4
·28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
·29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
·30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
·31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
·32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
·33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
·34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
·35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2286	1	1	2	2	3	3	4	4	5
·36	2291	2296	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
·37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
·38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
·39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
·40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
·41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
·42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685	1	1	2	2	3	4	4	5	6
·43	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748	1	1	2	3	3	4	4	5	6
·44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
·45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
·46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
·47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
·48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	4	4	5	6	6
·49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155	1	1	2	3	4	4	5	6	6

											Mean Differences								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
·50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
·51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304	1	2	2	3	4	5	5	6	7
·52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
·53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459	1	2	2	3	4	5	6	6	7
·54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540	1	2	2	3	4	5	6	6	7
·55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	7	7
·56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
·57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793	1	2	3	3	4	5	6	7	8
·58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
·59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	5	5	6	7	8
·60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064	1	2	3	4	5	6	6	7	8
·61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
·62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
·63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
·64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
·65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
·66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
·67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
·68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
·69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000	1	2	3	5	6	7	8	9	10
·70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	4	5	6	7	8	9	11
·71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
·72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
·73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	3	4	5	6	8	9	10	11
·74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
·75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
·76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
·77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
·78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
·79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
·80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
·81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
·82	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745	2	3	5	6	8	9	11	12	14
·83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
·84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063	2	3	5	6	8	10	11	13	15
·85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
·86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	8	10	12	13	15
·87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16
·88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
·89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	13	14	16
·90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
·91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
·92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
·93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	3	8	10	12	14	16	18
·94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
·95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
·96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
·97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
·98	9550	9572	9594	9616	9638	9661	9683	[illegible]	9727	9750	2	4	7	9	11	13	16	18	20
·99	9772	9795	9817	9840	9863	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20

Internal Assessment (Mid Examinations)

The pattern of Mid Exams or Continuous Internal Evaluation (CIE) prescribed by the JNTU-H as per the Regulations 2019 (R19) for all the semesters is as follows,

➢ There would be two Mid Exams or Continuous Internal Evaluation (CIE) for each semester,

- The I^{st} **Mid Term Examinations** would be conducted during the Middle of the Semester.
- The II^{nd} **Mid Term Examinations** during the last week of instructions.

➢ The Mid Exam I and II would have the same pattern of question paper which would carry **25 Marks** each and the time duration for conducting each Mid exam would be 120 min.

➢ The pattern of Mid Exam Question Paper would consist of two parts i.e., **Part-A** and **Part-B.**

- **Part-A** consist of 5 compulsory questions each carries 2 marks (i.e $5 \times 2 = 10$ marks).
- **Part-B** consist of 5 questions out of which 3 questions should be answered, each question carries 5 marks (i.e $5 \times 3 = 15$ marks).

➢ The average of the two Mid exams will be added with the 75 marks of External end examination which equals to 100 marks (i.e $25 + 75 = 100$).

UNIT - I

Part - A

1.	Define the term variable.	(Refer Unit-I, SQA-3)
2.	Define measurement.	(Refer Unit-I, SQA-5)
3.	Research	(Refer Unit-I, SQA-1)
4.	Compare and contrast between independent and dependent variable.	(Refer Unit-I, SQA-4)
5.	Ethics in business research.	(Refer Unit-I, SQA-10)

Part - B

1.	Explain the classification of research.	(Refer Unit-I, Q.No. 4)
2.	Explain the steps involved in research process.	(Refer Unit-I, Q.No. 5)
3.	Define the term variable. Explain various types of variables.	(Refer Unit-I, Q.No. 7)
4.	Define measurement. What are the characteristics of measurement?	(Refer Unit-I, Q.No. 9)
5.	What are the various sources of errors in measurement.	(Refer Unit-I, Q.No. 16)

UNIT - II

Part - A

1.	Components of Research Design.	(Refer Unit-II, SQA-1)
2.	Cross-sectional	(Refer Unit-II, SQA-3)

3. Define Questionnaire. (Refer Unit-II, SQA-7)
4. Sampling (Refer Unit-II, SQA-9)
5. What are the differences between Probability and Non-Probability Sampling? (Refer Unit-II, SQA-10)

Part - B

1. Define Research Design. What are the components of Research Design? (Refer Unit-II, Q.No. 3)
2. Discuss in detail classification/types of research designs. (Refer Unit-II, Q.No. 7)
3. Discuss briefly action research and its process. (Refer Unit-II, Q.No. 9)
4. What are the Characteristics of the Good Research Design? (Refer Unit-II, Q.No. 10)
5. What are the sources and methods of collecting data? (Refer Unit-II, Q.No. 13)
6. What are the difference between Primary and Secondary Data ? (Refer Unit-II, Q.No. 20)
7. Explain different types of Sampling Methods. (Refer Unit-II, Q.No. 31)
8. What are the Applications of Sampling? (Refer Unit-II, Q.No. 33)

UNIT - III

Part - A

1. What is tabulation ? Explain the significance of tabulation. (Refer Unit-III, SQA-1)
2. Define classification. (Refer Unit-III, SQA-3)
3. Explain advantages and disadvantages of diagrams. (Refer Unit-III, SQA-6)
4. Bar Diagram. (Refer Unit-III, SQA-7)
5. Applications of t-distribution. (Refer Unit-III, SQA-8)
6. Explain briefly about Paired t-test. (Refer Unit-III, SQA-10)

Part - B

1. What is tabulation ? Explain the signifi- cance of tabulation. (Refer Unit-III, Q.No. 1)
2. Explain different parts of a table. (Refer Unit-III, Q.No. 2)
3. Explain different types of classification? (Refer Unit-III, Q.No. 8)
4. Define diagram and diagramatical represenation. Explain general rules for constructing diagrams. (Refer Unit-III, Q.No. 12)
5. What do you mean by one-dimensional diagram? What are the commonly used one-dimensional diagrams? (Refer Unit-III, Q.No. 17)
6. Explain different types of bar diagram. (Refer Unit-III, Q.No. 18)
7. Define Small sample tests and t-distribution. (Refer Unit-III, Q.No. 21)

UNIT - IV

Part - A

1. What is ANOVA? What are its assump-tions and applications? (Refer Unit-IV, SQA-1)
2. Chi-Square Distribution. (Refer Unit-IV, SQA-2)
3. Define scatter diagram. Explain merits and demerits of scatter diagram. (Refer Unit-IV, SQA-5)
4. Merits and demerits of coefficient of correlation. (Refer Unit-IV, SQA-6)
5. Multiple Regression Analysis (Refer Unit-IV, SQA-11)
6. Advantages of Factor Analysis (Refer Unit-IV, SQA-13)
7. Objectives of Cluster Analysis (Refer Unit-IV, SQA-15)

Part - B

1. What is ANOVA? What are its assump-tions and applications? (Refer Unit-IV, Q.No. 1)
2. Explain briefly about One Way ANOVA. (Refer Unit-IV, Q.No. 2)
3. Explain briefly about two way ANOVA with and without interation? (Refer Unit-IV, Q.No. 3)
4. Explain briefly about test for goodness of fit? (Refer Unit-IV, Q.No. 7)
5. Explain various types of Correlation. (Refer Unit-IV, Q.No. 10)
6. Define scatter diagram. Explain merits and demerits of scatter diagram. (Refer Unit-IV, Q.No. 14)
7. Define the term Regression. Explain the utility of Regression Analysis. (Refer Unit-IV, Q.No. 22)
8. Define Exploratory Factor Analysis. What are the objectives of Exploratory Factor Analysis? (Refer Unit-IV, Q.No. 36)
9. Define conjoint analysis. What are the objectives of conjoint analysis. (Refer Unit-IV, Q.No. 50)

UNIT - V

Part - A

1. Define Time Series. (Refer Unit-V, SQA-1)
2. Define trend analysis. (Refer Unit-V, SQA-2)
3. What is free hand curve? (Refer Unit-V, SQA-3)
4. Index numbers (Refer Unit-V, SQA-5)
5. Uses of index numbers. (Refer Unit-V, SQA-6)
6. Oral Presentation (Refer Unit-V, SQA-9)
7. Referencing styles (Refer Unit-V, SQA-10)

Part - B

1. Explain the Utility of Time Series Analysis. (Refer Unit-V, Q.No. 2)
2. What are the Components of Time Series ? (Refer Unit-V, Q.No. 3)
3. Define Trend Analysis. Explain the purpose of measuring trend. (Refer Unit-V, Q.No. 5)

4. Define least square method. Explain merits and demerits of least square method. (Refer Unit-V, Q.No. 11)
5. Explain the Characteristics of Index Numbers. (Refer Unit-V, Q.No. 13)
6. Explain the uses of index numbers. (Refer Unit-V, Q.No. 14)
7. Discuss in detail consumer price index numbers/cost of living index numbers. (Refer Unit-V, Q.No. 24)
8. State the Guides for Effective Documentation. (Refer Unit-V, Q.No. 32)

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

M.B.A I Year I Semester Examination **R19**

MODEL PAPER - I

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 3 Hours] [Max. Marks : 75

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

PART - A (5 × 5 = 25 Marks)

ANSWERS

1. (a) Characteristics of Research **(Unit-I, SQA-2)**

(b) What are the differences between Probability and Non-Probability Sampling? **(Unit-II, SQA-10)**

(c) What are the differences between diagrams and graphs? **(Unit-III, SQA-11)**

(d) Define multiple regression analysis. **(Unit-IV, SQA-11)**

(e) Calculate the 3-yearly moving averages of the production figures given below and draw the trend:

Year	Production (in met. tonnes)	Year	Production (in met. tonnes)
1985	15	1993	63
1986	21	1994	70
1987	30	1995	74
1988	36	1996	82
1989	42	1997	90
1990	46	1998	95
1991	50	1999	102
1992	56	–	–

(Unit-V, Prob. 1)

PART - B (5 × 10 = 50 Marks)

2. Define research ? Explain different types of research. **(Unit-I, Q.No. 1, 4)**

(OR)

3. Define measurement. What are the characteristics of measurement? **(Unit-I,Q.No.9)**

4. Define Research Design. What are the components of Research Design? **(Unit-II, Q.No.3)**

(OR)

5. (a) Define Questionnaire. Explain different types of Questionnaire. **(Unit-II, Q.No.21)**

(b) What are the advantages of sampling ? **(Unit-II, Q.No.29)**

6. Explain different types of classification? **(Unit-III, Q.No.8)**

(OR)

7. A manufacturer of certain electric Bulbs claims that is bulbs have mean life 25 months with standard deviation 5 months a random sample of 6 such bulbs areas follows.

Life of Bulbs	**1**	**2**	**3**	**4**	**5**	**6**
Months	**24**	**26**	**30**	**20**	**20**	**18**

Can you regard the procedure clamied to be valid at 1% Los ($t_{0.01} = 4.032$). **(Unit-III, Prob.9)**

8. (a) A die is rolled 100 times with the following distribution:

Number	1	2	3	4	5	6
Observed Frequency	17	14	20	17	17	15

At the 0.01 level of significance, determine whether die is true (or uniform). **(Unit-IV, Prob.11)**

8. (b) Explain briefly about test for goodness of fit. **(Unit-IV, Q.No. 4)**

(OR)

9. (a) From the following data, calculate the regression equations taking deviation of items from the mean of X and Y series.

X	6	2	10	4	8
Y	9	11	5	8	7

(Unit-IV, Prob.27)

(b) What is discriminant analysis? **(Unit-IV, Q.No. 31)**

10. Given below is the data of production of a certain company in lakhs of units.

Year	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**
Production	**15**	**14**	**18**	**20**	**17**	**24**	**27**

Compute the Linear trend by the method of least squares. predict the production for the year “2004”.

(Unit-V, Prob.6)

(OR)

11. Calculate Fisher’s price index from the following data and check whether Time Reversal Test is satisfied :

Commodity	Base Year		Current Year	
	Price (Rs.)	Qty. (Kg.)	Price (Rs.)	Qty (Kg.)
A	32	50	30	50
B	30	35	25	40
C	16	55	18	50

(Unit-V, Prob.19)

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

M.B.A I Year I Semester Examination **R19**

MODEL PAPER - II

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 3 Hours] [Max. Marks : 75

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

PART - A (5 × 5 = 25 Marks)

ANSWERS

1. (a) Ordinal Scale **(Unit-I, SQA-8)**

(b) Disadvantages of Secondary Data **(Unit-II, SQA-6)**

(c) Represent the below given data by a simple bar diagram :

Year	Sales Tax (in '000)
1961	900
1966	1800
1971	2500
1976	4500
1981	6000

(Unit-III, Prob. 7)

(d) What is discriminant analysis. **(Unit-IV, SQA-12)**

(e) From the following data, construct an Index number for 1994 taking 1999 as base as per simple Aggregative method

Commodities	Price in 1993 (Rs)	Price in 1994 (Rs)
A	40	60
B	60	90
C	85	125
D	25	35
E	30	40

(Unit-V, Prob. 12)

PART - B (5 × 10 = 50 Marks)

(Essay Type Questions)

2. Explain the steps involved in research process. **(Unit-I, Q.No.5)**

(OR)

3. What do you understand by ethics in business research? **(Unit-I,Q.No.17)**

4. Explain different types of Sampling Methods. **(Unit-II, Q.No.31)**

(OR)

5. What are the difference between Primary and Secondary Data ? **(Unit-II, Q.No.20)**

6. Draw a pie diagram for the following data of Sixth Five - Year Plan Public Sector outlays :

Agriculture and Rural Development	12.9%
Irrigation, etc.	12.5%
Energy	27.2%
Industry and Minerals	15.4%
Transport, Communication, etc.	15.9%
Social Services and Others	16.1%

(Unit-III, Prob.4)

(OR)

7. The life a time of electric bulbs for a random sample of 10 from a large consignment given the following data.

Item	1	2	3	4	5	6	7	8	9	10
Life in hours	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

We can accept the Hypothesis that Average life time of is 4 hours test at 5% level of significance.

(Unit-III, Prob.11)

8. (a) Define Exploratory Factor Analysis. What are the objectives of Exploratory Factor Analysis?

(Unit-IV, Q.No. 36)

(b) What are the merits and demerits of rank correlation ? **(Unit-IV, Q.No. 21)**

(OR)

9. Four machines A, B, C and D are used to produce a certain kind of cotton fabrics. Samples of size 4 with each unit as 100 square metres are selected from the outputs of the machines at random and the number of flaws in each 100 square metres show the following result.

A	B	C	D
8	6	14	20
9	8	12	22
11	10	18	25
12	4	9	23

Do you think that there is a significant difference in the performance of the four machines?

(Unit-IV, Prob.2)

10. (a) What are the Components of Time Series ? **(Unit-V, Q.No. 3)**

(b) Discuss the referencing styles and citation in business management research.

(Unit-V, Q.No. 34)

(OR)

11. From the following data, calculate Price Index numbers by

(a) Laspeyre's

(b) Paasche's and

(c) Fisher's ideal method.

Commodity	2005		2009	
	Price `	Qty.	Price `	Qty.
A	10	4	20	3
B	25	5	30	3
C	20	8	25	8
D	10	10	10	13

(Unit-V, Prob.17)

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

M.B.A I Year I Semester Examination **R19**

MODEL PAPER - III

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 3 Hours] [Max. Marks : 75

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

PART - A (5 × 5 = 25 Marks)

ANSWERS

1. (a) Ethics in business research **(Unit-I, SQA-10)**

(b) What are the Requirements for a Good Research Design? **(Unit-II, SQA-2)**

(c) What are the applications of t-distribution. **(Unit-III, SQA-8)**

(d) Find the coefficient of correlation between x and y

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

(Unit-IV, Prob. 14)

(e) Explain the uses of index numbers. **(Unit-V, SQA-6)**

PART - B (5 × 10 = 50 Marks)

(Essay Type Questions)

2. Define the term variable. Explain various types of variables. **(Unit-I, Q.No. 7)**

(OR)

3. Explain the test of reliability and validity of measurement. **(Unit-I,Q.No.14, 15)**

4. (a) What is Research Problem? Explain the various conditions involved in research problem. **(Unit-II, Q.No.1)**

(b) What are the Applications of Sampling? **(Unit-II, Q.No.33)**

(OR)

5. What are the Guidelines / Precautions for Preparation of Questionnaire? **(Unit-II, Q.No.24)**

6. Define tabulation ? Explain general rules for tabulation. **(Unit-III, Q.No.1, 3)**

(OR)

7. Eleven sales executive trainees are assigned selling jobs right after their recruitment. After a fortnight they are withdrawn from their field duties and given a month's training for executives sales. Sales executed by them in thousands of rupees before and after the training in the same period are listed below,

Sales (000 ₹) (Before training)	**23**	**20**	**19**	**21**	**18**	**20**	**18**	**17**	**23**	**16**	**19**
Sales (000 ₹) (After training)	**24**	**19**	**21**	**18**	**20**	**22**	**20**	**20**	**23**	**20**	**27**

Do these data indicate that the training has contributed to their performance?

(Unit-III, Prob.13)

8. Calculate Karl Pearson's Coefficient of correlation between expenditure on advertising and sales from the data given below

Advertising Expenses ('000 Rs.)	39	65	62	90	82	75	25	98	36	78
Sales (Lakh Rs.)	47	53	58	86	62	68	60	91	51	84

(Unit-IV, Prob.13)

(OR)

9. Four different drugs have been developed for a certain disease. These drugs are used under three different environments. It is assumed that the environment might affect efficacy of drugs. The number of cases of recovery from the disease per 100 people who have taken the drugs is tabulated as follows :

Environment	Drug A1	Drug A2	Drug A3	Drug A4
I	19	8	23	8
II	10	9	12	6
III	11	10	13	16

Test whether the drugs differ in their efficacy to treat the disease, also whether there is any effect of environment on the efficacy of disease.

(Unit-IV, Prob.3)

10. (a) Discuss various types of Research Reports. **(Unit-V, Q.No. 27)**

(b) Describe the structure of an effective report. **(Unit-V, Q.No. 30)**

(OR)

11. The following data relate to the number of passenger cars (in millions) sold from 2004 to 2011.

Years	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**
Number	**6.7**	**5.3**	**4.3**	**6.1**	**5.6**	**7.9**	**5.8**	**6.1**

(a) Fit a straight line trend to the data through 2009.

(b) Estimate production in 2011 and compare with the actual production.

(Unit-V, Prob.11)

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

MBA I - Semester Examinations

R19

October / November - 2021

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 2 Hours | Max. Marks : 75

Answer any five questions
All questions carry equal marks

Answers

1. (a) What are the challenges faced in conduct of social science research ?

Ans :

Research in social sciences has certain limitations and problems when compared with research in physical sciences. They are discussed below:

(a) Scientists a part of what is studied: The fact that social scientist is part of the human society which he studies gives rise to certain limitations. Man must have to be his won guinea pig. as pointed out by Jalian Huxley. This has a number of methodological consequences. For example, it restricts the scope for controlled experiments. It limits the scope for objectivity in social science research.

(b) Complexity of the subject matter: The subject matter of research in social science, viz. human society and human behaviour is too complex varied and changing to yield to the scientific categorization, measurement, analysis and prediction. The multiplicity and complexity of causation make it difficult to apply the technique of experimentation. Human behaviour can be studied only be other human beings, and this always distorts fundamentally the facts being studied so that there can be no objective procedure for achieving the truth.

(c) Human Problems: A social scientist faces certain human problems, which the natural scientist is sparred. These problems are varied and include refusal of respondents improper understanding of questions by them their loss of memory, their reluctance to furnish certain information, etc. All these problems cause biases and invalidate the research findings and conclusions.

(d) Personal Values: Subjects and clients, as well as investigators, have personal values, which are apt to become involved in the research process. One should not assume that these are freely exploitable. The investigator must have respect for the client's values.

(e) Anthropomorphization: Another hazard of social science research is the danger of the temptation to anthropomorphize about humans, it results in using observation obtained by sheer intuition or empathy in conceptualizing in anthropomorphic manner.

(f) Wrong Decisions: The quality of research findings depends upon the soundness of decisions made by the social scientist on such crucial stages of his research process as definition of the unit of study operationalization of concepts, selection of sampling techniques and statistical techniques. Any mistake in any of these decisions will vitiate the validity of his findings.

(b) What are the ethical issues concerning research participants ? **(Unit-I, Q.No. 18)**

2. (a) What is a diagnostic study ? What is the purpose of it ?

Ans :

- Diagnostic refers to scientific differentiation among various conditions or phenomenon for the purpose of accurately classifying these conditions.
- In its broadest sense diagnosis corresponds to the fact-finding aspects of clinical practice.
- Its objective includes screening and classification personality description, prediction of outcome and attainment of insight by the client.
- The diagnostic research paradigm represents the most typical and simple problem solving strategy of the helper faced with problems and crises on the job.
- It consists of:
  (a) the emergence of a problem
  (b) a diagnosis of its causes
  (c) formulation of all possible avenues of remediation, and
  (d) recommendations for a possible solution.

(b) Distinguish between questionnaire and schedule for collection of data.

Ans :

BASIS FOR COMPARISON	QUESTIONNAIRE	SCHEDULE
Meaning	Questionnaire refers to a technique of data collection which consist of a series of written questions along with alternative answers.	Schedule is a formalized set of questions, statements and spaces for answers, provided to the enumerators who ask questions to the respondents and note down the answers.
Filled by	Respondents	Enumerators
Response Rate	Low	High
Coverage	Large	Comparatively small
Cost	Economical	Expensive
Respondent's identity	Not known	Known
Success relies on	Quality of the questionnaire	Honesty and competence of the enumerator.
Usage	Only when the people are literate and cooperative.	Used on both literate and illiterate people.

3. What is secondary source of data ? Critically evaluate its use for research. What are the limitations? **(Unit-II, Q.No. 17)**

4. (a) What is multivariate table ? Illustrate your answer with a specimen copy of table containing fictitious data. **(Unit-III, Q.No. 6)**

   (b) Under what conditions can we apply 't' test for test of hypothesis ? **(Unit-III, Q.No. 21, 22)**

5. A random sample of 10 boys had the following IQs :

   70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean IQ of 100? Find a reasonable range in which most of the mean IQ values of sample of 10 boys lie.

Sol :

There standard deviation of mean of the sample is not given directly.

So we have to determine these S.D X mean as follows :

$$\text{Mean } \bar{X} = \frac{\Sigma X}{h} = \frac{972}{10} = 97.2$$

X	**X – $\bar{X}$**	**(X – $\bar{X}$)²**
70	–27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	–9.2	84.64
83	–14.2	201.64
95	–2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
	0	1833.60

We know that $S^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2 = \frac{1833.60}{9}$

$\therefore$ Standard Deviation, $S = \sqrt{203.73} = 14.27$

(i) Null Hypothesis (H_0) : The daa support the assumption of a population mean I.Q of 100 in the population

(ii) Alternative Hypothesis (H_1) : $\mu \neq 100$

(iii) Level of Significance (α) : 0.05

(iv) Test Statistics $t_{cal} = \dfrac{\bar{X}-\mu}{S/\sqrt{n}} = \dfrac{97.2-100}{14.27/\sqrt{10}} = -0.62$

$\therefore$ $|t| = 0.62$ i.e., calculated value of t = 0.62

(v) Conclusion : Tabulated value of t f_0 (10 – 1) = 9 d.f

i.e., 9 d.f at 5% L.O.S 2.26 (Two tailed test)

$\because$ Cal value of t < Tabulated value of t, we accept the null hypothesis H_0.

i.e., The data support the assumption of mean I.Q of 100 in the population

* The 95% confidence limits are given by $\bar{X} \pm t_{0.05} \, S/\sqrt{n}$

$= 9.72 \pm 2.26 \pm 4.512 = 9.72 \pm 10.198 = [107.4, 8.7]$

$\therefore$ The 95% confidence limits within which the mean I.Q values of samples of 10 boys will lie is [9.7, 107]

6. Two housewives, Reena and Reshma asked to expenses their preference for different kinds of detergents, gave the following replies :

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

To what extent the preference of these ladies go together ?

Sol :

(i) Null Hypothesis (H_0) : There is no significance different between the preference of these ladies go together.

(ii) Alternative Hypothesis (H_1) : There is significance different between the preference of these ladies go together.

(iii) Level of Significance (α) : 0.05 (Assume)

(iv) Calculation for sample means × Standard Deviation's

Here $n_1 = 10$, $n_2 = 10$

$\therefore \quad \bar{X} = \dfrac{\Sigma X}{n} = \dfrac{55}{10} = 5.5 \; ; \; \bar{Y} = \dfrac{\Sigma Y}{n} = \dfrac{55}{10} = 5.5$

X	Y	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$
1	1	20.25	20.25
2	4	12.25	2.25
4	2	2.25	12.25
3	3	6.25	6.25
7	5	6.25	0.25
8	7	6.25	2.25
6	6	0.25	0.25
5	8	0.25	6.25
9	9	12.25	12.25
10	10	20.25	20.25
55	55	68.5	68.5

$$\frac{\sum_{i-1} (X_i - \bar{X})^2 + \sum_{i=1} (Y_i - \bar{Y})^2}{n_1 + n_2 - 2} = \frac{68.5 + 68.5}{10 + 10 - 2}$$

S.D = 7.611

(v) Test Statistics $t = \dfrac{\bar{X} - \bar{Y}}{\sqrt{S^2\left[\dfrac{1}{n_1} + \dfrac{1}{n^2}\right]}} = \dfrac{5.5 - 5.5}{\sqrt{57.92\left(\dfrac{1}{10} + \dfrac{1}{10}\right)}}$

$$\frac{0}{3.403} = 0.$$

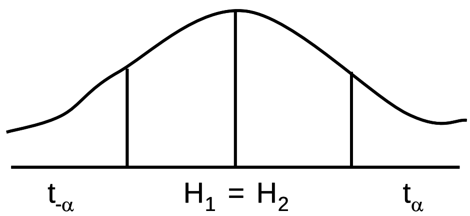
Degrees of freedom = $n_1 + n_2 - 2 = 10 + 10 - 2 = 18$

Tabulated t for 18 d.f at 5% level = 1.734

$\because$ Cal t < tab t

$\therefore$ We accept the null hypothesis

H_0 . The two ladies asked to express this preference different kinds of detergents are different



7. Three different methods of teaching statistics are used on three groups of students, Random samples of size 5 are taken from each group and the results are shown below.

The grades are on a 10 point scale :

Group A	Group B	Group C
7	3	4
6	6	7
7	5	7
7	4	4
8	7	8

Determine on the basis of the above data whether there is a difference in the teaching methods.

Sol :

(i) Null Hypothesis (H_0) : There is no difference between in the teaching methods.

(ii) Alternative Hypothesis (H_1) : There is a difference between in the teaching methods

(iii) Level of Significance (α) : 0.05 (Assume)

(iv) Calculations

Group 'A'	**Group 'B'**	**Group 'C'**
7	3	4
6	6	7
7	5	7
7	4	4
8	7	8
$T_1 \rightarrow 35$	$T_2 \rightarrow 35$	$T_3 \rightarrow 30$

Grand Total = 35 + 25 + 30 = 90

(i) Correction factor (c.f) = $\frac{(GT)^2}{N} = \frac{(90)^2}{15} = \frac{8100}{15} = 540$

(ii) Total sum of square (TSS) = $\Sigma\,(X_{ij}^2) - C.F = 576 - 540$

TSS = 36

(iii) Sum of square between samples

$$\frac{\Sigma T_j^2}{n_j} - C.F$$

$$\left[\frac{(35)^2}{5} + \frac{(25)^2}{5} + \frac{(30)^2}{5}\right] - 540$$

550 – 540

$\boxed{SSB = 10}$

(iv) Sum of the squares of errors

SSE = TSS – SSR

36 – 10

$\boxed{SSE = 26}$

ANOVA One Way Table

Source of Variation	Degree of Freedom	Sum of Square	Mean Sum of Square	F – Ratio & $F_{calculation}$
SSB	3 – 1 = 2	10	$10/2 = 5$	
SSE	15 – 3 = 12	26	$26/12 = 2.166$	$F_{cal} = \frac{5}{2.166} = 2.303$
TSS	15 – 1 = 14	36	–	

Conclusion :

F_{cal} & F_{Ratio} = 2.166

F_{tab} at (2, 12) degree of freedom F_{tab} = 3.74

$F_{tab} < F_{tab}$

$\therefore$ H_0 is accepted

We conclude that there is no difference in the teaching method.

8. The materials manager of a company has projected 10, 15 and 18 trackload of product for three consecutive months. The seasonal indices for these are 141.5, 128.5 and 82.6 respectively. Work out the seasonalized forecast for each month of three months.

Sol :

Month	Truck Load	Seasonal Indices
1	10	141.5
2	15	128.5
3	18	82.6

Seasonal forecast of 1st month = $\frac{10}{141.5} \times 128.5 = 9.07$

Seasonal forecast of 2nd month = $\frac{15}{128.5} \times 82.6 = 9.64$

Seasonal forecast of 3nd month = $\frac{18}{82.6} \times 141.5 = 30.83$

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

MBA I - Semester Examinations

R19

July / August - 2021

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 2 Hours Max. Marks : 75

Answer any five questions
All questions carry equal marks

Answers

1. (a) What are the limitations of social science research ? **(Oct./Nov.-21, Q.No. 1(a))**

   (b) What are the levels of measurement ? **(Unit-I, Q.No. 13)**

2. (a) What are the characteristics of good measurement tool ? **(Unit-I, Q.No. 9)**

   (b) What is meant by exploratory research? What are the purposes for which it is conducted?

Ans :

Exploratory research, as the name implies, intends merely to explore the research questions and does not intend to offer final and conclusive solutions to existing problems. This type of research is usually conducted to study a problem that has not been clearly defined yet.

Conducted in order to determine the nature of the problem, exploratory research is not intended to provide conclusive evidence, but helps us to have a better understanding of the problem. When conducting exploratory research, the researcher ought to be willing to change his/her direction as a result of revelation of new data and new insights

3. (a) What are the sources of research problem ?

Ans :

(i) Strategic Level Research Problems

The strategic level pertains to problems encountered by top-level managers such as looking into the possibility of merging with another company. Research can provide the information needed to objectively examine the pros and cons of such endeavor.

(ii) Executory Management Research Problems

Problems of execution through middle-level managers can serve as the focus of the study. Since this level of management lies between the top-level management and the operational-level management, issues related to the operational manager's performance of assigned tasks aligned with the organization's goals serve as research areas. For instance, if the performance of duties is below par, the effect of specific incentives such as generous overtime pay to supervisors may be explored.

(iii) Operational Management Research Problems

At the operational management level, workers or line and staff job behavior and performance are rich sources of research problems. For example, a researcher can focus his attention on the effect of strict punctuality policy on employee productivity.

(b) Explain "Before-after experimental design with control group".

Ans :

The inclusion of one control group in this design is aimed at taking account of the effects both of the initial measurement and of contemporaneous, external factors. In such a design, experimental and the control group are both measured at the beginning and also at the end of the experimental period.

Conditiions	**Experimental Groups**	**Control Group I**	**Control Group II**
Before Measurement	Yes (Y_1)	Yes (Y_1)	No (Y_1) = $(Y_1 + Y_2)/2$
Exposure to experimental factor	Yes	No	Yes
Exposure to uncontrolled events	Yes	Yes	Yes
After Measurement	Yes (Y_2)	Yes (Y_2)	Yes (Y_2)
Change	$d_1 = (Y_2 - Y_1)$	$d_2 = (Y_2 - Y_1)$	$d_3 = (Y_2 - Y_1)$

Interaction = $d_1 - (d_2 + d_3)$

This design involves an addition of one more control group to the previous design, i.e., 'Before-After' study with one control- group. This second control group is not pre- measured but is exposed to the experimental variable and subjected, of course, to after measurement.

The 'before' measure of the second control group is assumed to be similar to the 'before' measures of the experimental and the first control group, i.e., equal to the average of the 'before' measure of the experimental group and control group I. Thus, in control group II, there is exposure to experimental variable but no possibility of interaction between the 'before' measure and the experimental variable.

4. (a) Why is graphic representation of data desirable ? **(Unit-III, Q.No. 14)**

(b) A machine designed to produce insulating washers for electrical devices of average machines of 0.025 cm. A random sample of 10 washers was found to have an average thickness of 0.024 cm with a standard deviation of 0.002 cm. Test the significance of the deviation (value of t for 9 degrees of freedom at 5% level is 2.262).

Ans :

n = 10

n < 10

$\therefore$ The given sample is small sample (t - test)

Sample mean ($\bar{x}$) = 0.024 cm

Population mean (μ) = 0.025 cm

Population S.D (σ) = 0.002 cm

Sample size (η) = 10

(i) Null Hypothesis (H_0) : $\mu = 0.025$

(ii) Alternative Hypothesis (H_1) : $\mu \neq 0.025$

(iii) Level of significance (α) : 0.05

(iv) Test statistics $t_{cal} = \dfrac{\bar{X}-\mu}{S/\sqrt{n-1}} - \dfrac{0.024-0.025}{0.002/\sqrt{10-1}}$

$$\frac{-0.001}{0.000666} = -1.515$$

$|t| = |-1.515| = 1.515$

(v) Conclusion : t_{cal} 1.515

t_{tab} at 9 degree of freedom 5% level is 2.262

$t_{cal} < t_{tab}$

H_0 is accepted

5. Two researchers adopted different sampling techniques while investigating the same group of students to find the number of students falling in different intelligence levels. Can we say that the sampling techniques adopted by the two researchers are significantly different? Results of the investigation are shown below :

Researcher	Number of students in each level			
	Below Average	Average	Above Average	Genius
X	86	60	44	10
Y	40	33	25	2

Sol :

(i) Null Hypothesis (H_0) : There is no significance difference between the two researchers.

(ii) Alternative Hypothesis (H_1) : There is a significance difference between the two researchers.

(iii) Level of significance (α) : 0.05 (Assume).

(iv) Calculations :

Table of Expected Frequencies

Researcher	Below Average	Average	Above Average	Genius	Total
X	$\frac{126\times200}{300}=84$	$\frac{93\times200}{300}=62$	$\frac{69\times200}{300}=46$	$\frac{12\times200}{300}=8$	200
Y	$\frac{126\times100}{300}=42$	$\frac{93\times100}{300}=31$	$\frac{69\times100}{300}=23$	$\frac{12\times100}{300}=4$	100
	126	93	69	12	300

Calculation of χ^2

Observed Frequency O_i	Expected Frequency E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
86	84	4	0.0476
60	62	4	0.0645
44	46	4	0.0869
10	8	4	0.5
40	42	4	0.0952
33	31	4	0.129
25	23	4	0.1739
2	4	4	1
			2.0941

(v) New $\chi^2_{Cal} = \Sigma \frac{(O_i - E_i)^2}{E_i} = 2.0971$

i.e., Calculated $\chi^2 = 2.0971$

Conclusion

Tabulated χ^2 for (2 – 1) (4 – 1) = 3 d.f. at 5% L.O.S is 7.92.

$\because$ Calculated χ^2 < tabulated χ^2, We accept the null hypothesis H_0.

i.e., there is no significance difference between the two researchers.

6. (a) What is cluster analysis ? What are its applications ? **(Unit-IV, Q.No. 45)**

Ans :

Applications

Cluster analysis is used in many applications including pattern recognition, marketing research, image processing and data analysis. It can also help marketers and influencers to discover target groups as their customer base. After that, it can characterize these groups based on a customer's purchasing patterns. It also helps with data presentation and analysis.

Clustering analysis also helps in the field of biology. It can be used to categorize genes, derive animal and plant taxonomies and can gain insights into various structures and inherent population patterns.

It can also help in the detection of land use in an area. One can also identify groups of houses in a particular city based on geographical location, house type, value etc. Clustering also helps to organize and classify documents for certain information discovery. Clustering can also be used in detection and protection from fraud activities such as detection of credit card fraud.

(b) Explain clearly the differences between simple linear and multiple linear regressions.

Ans :

Simple Regression	**Multiple Regression**
1. Dependent variable Y predicted from one independent variable X	1. Dependent variable Y predicted from a set of independent variables (X_1, X_2 ... X_k)
2. Regression coefficient	2. Regression coefficient for each independent variable.
3. r^2 proportion of variation in dependent variable Y predictable from X.	3. R^2 proportion of variable in dependent variable Y predictable by set of independent variables (X's)

7. Fit a straight line trend to the following time series data :

Year	2015	2016	2017	2018	2019
Sale of Sugar in thousand kg.	80	90	92	83	94

Sol :

Year	**Sales (Y)**	**X**	**X^2**	**XY**
2015	80	–2	4	–160
2016	90	–1	1	–90
2017	92	0	0	0
2018	83	1	1	83
2019	94	2	4	188
	439		10	21

$$y_c = a + bX$$

$$a = \frac{\Sigma y}{n} = \frac{439}{5} - 87.5$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{21}{10} = 2.1$$

$$y_c = 87.8 + 2.1\,(X)$$

2015 = 87.8 + 2.1 (–2) = 87.8 – 4.2 = 83.6

2016 = 87.8 + 2.1 (–1) = 87.8 – 2.1 = 85.7

2017 = 87.8 + 2.1 (0) = 87.8 + 0 = 87.8

2018 = 87.8 + 2.1 (1) = 87.8 + 2.1 = 89.9

2019 = 87.8 + 2.1 (2) = 87.8 + 4.2 = 92

8. (a) What are the purpose of footnotes? What is the kind of footnotes one finds in research.

Ans :

Footnotes are notes placed at the bottom of a page. They cite references or comment on a designated part of the text above it. For example, say you want to add an interesting comment to a sentence you have written, but the comment is not directly related to the argument of your paragraph. In this case, you could add the symbol for a footnote.

When your reader comes across the footnote in the main text of your paper, he or she could look down at your comments right away, or else continue reading the paragraph and read your comments at the end. Because this makes it convenient for your reader, most citation styles require that you use either footnotes or endnotes in your paper. Some, however, allow you to make parenthetical references (author, date) in the body of your work. See our section on citation styles for more information.

Footnotes are not just for interesting comments, however. Sometimes they simply refer to relevant sources — they let your reader know where certain material came from, or where they can look for other sources on the subject. To decide whether you should cite your sources in footnotes or in the body of your paper, you should ask your instructor or see our section on citation styles.

(b) What is an index number? Describe briefly in application in business and industry. **(Unit-V, Q.No. 12)**

Ans :

Applications

1. In Measuring Changes in the Value of Money

Index numbers are used to measure changes in the value of money. A study of the rise or fall in the value of money is essential for determining the direction of production and employment to facilitate future payments and to know changes in the real income of different groups of people at different places and times.

2. In Cost of Living

Cost of living index numbers in the case of different groups of workers throw light on the rise or fall in the real income of workers. It is on the basis of the study of the cost of living index that money wages are determined and dearness and other allowances are granted to workers. The cost of living index is also the basis of wage negotiations and wage contracts.

3. In Analyzing Markets for Goods and Services

Consumer price index numbers are used in analysing markets for particular kinds of goods and services. The weights assigned to different commodities like food, clothing, fuel, and lighting, house rent, etc., govern the market for such goods and services.

4. In Measuring Changes in Industrial Production

Index numbers of industrial production measure increase or decrease in industrial production in a given year as compared to the base year. We can know from such as index number the actual condition of different industries, whether production is increasing or decreasing in them, for an industrial index number measures changes in the quantity of production.

5. In Internal Trade

The study of indices of the wholesale prices of consumer and industrial goods and of industrial production helps commerce and industry in expanding or decreasing internal trade.

6. In External Trade

The foreign trade position of a country can be accessed on the basis of its export and import indices. These indices reveal whether the external trade of the country is increasing or decreasing.

7. In Economic Policies

Index numbers are helpful to the state in formulating and adopting appropriate economic policies. Index numbers measure changes in such magnitudes as prices, incomes, wages, production, employment, products, exports, imports, etc. By comparing the index numbers of these magnitudes for different periods, the government can know the present trend of economic activity and accordingly adopt price policy, foreign trade policy and general economic policies.

8. In Determining the Foreign Exchange Rate

Index numbers of wholesale price of two countries are used to determine their rate of foreign exchange. They are the basis of the purchasing power parity theory which determines the exchange rate between two countries on inconvertible paper standard.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

MBA I-Semester Examinations

October / November - 2020

R19

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 2 Hours] [Max. Marks : 75

Answer any Five questions

All questions carry equal marks

ANSWERS

1. (a) What is the importance of measurement in research ? What are the common types of scales used for measurement in social science ? **(Unit-I, Q.No.9,10)**

(b) What are the unique challenges in conduct of social science research ?

Ans :

Social research is at cross-roads in developing countries like India. While the country has the highest volume of research in the region, and is significantly ahead of other countries in south Asia, there is wide disparity in research activity and output across the country, both in terms of quantity and quality. Some premium universities located in the major cities foster academic research cultures which include interdisciplinary work, knowledge production with emphasis on peer review, and engagement with internal and external intellectual networks and learned societies

(i) In India, social problems are many more and much complex, as compared with developed countries.

(ii) Country like India cannot spare funds for social research because needs of money in other walks of life are much pressing and require immediate attention than research.

(iii) It is not easy to raise additional resources for research because a vicious circle of poverty, capital formation, taxation etc., is going on.

(iv) Most of the researchers in India are not well equipped with latest research techniques.

(v) Universities and research organizations in India practically have no infrastructure to produce trained and qualified social researchers, at short notice.

(vi) The number of qualified, trained and devoted social research workers is already much less than what is needed to investigate and research social problems. Etc.

2. (a) Who are the stakeholders in research ? Who are the participants in research?

Ans :

Stakeholders are people or organizations who have an interest in your research project, or affect or are affected by its outcomes. Stakeholders include those who are both supportive of your research, as well as those who may be less supportive or indeed critical of it.

The purpose of stakeholder analysis is to :

- Identify project stakeholders
- Determine what interest each stakeholder has in your project
- Assess how much influence stakeholders have on the project
- Consider how you will manage and communicate with different types of stakeholder.

(b) Distinguish between the terms ‘reliability’ and ‘validity’.

Ans :

S.No.	Validity	S.No.	Reliability
1.	Validity refers to a situation when a test (or) instrument is accurately measuring what it’s supposed to.	1.	Reliability refers to the degree of reproducibility of the results if the measurement is repeated.
2.	A valid instrument is always reliable	2.	A reliable instrument is not valid
3.	Validity is important while evaluating the multi-item scale.	3.	Reliability has no role to play while evaluating a multi-item scale.
4.	Assessing validity is a difficult task	4.	Assessing reliability is easy
5.	Validity checks whether the scale produces the expected result.	5.	Reliability concentrates on precision, which measures the extent to which a scale produces a consistent outcome.

3. What is ‘experimental design’ in social science research ? What is the relevance of “Control Group” in experiment ? Explain any four formats of ‘validity’. **(Unit-II, Q.No.8)**

4. (a) What are the advantages and disadvantages associated with cross-tabulation as a method of presenting and analysis of data ? **(Unit-II, Q.No.7)**

Ans :

Advantages and disadvantages associated with cross tabulation are :

Advantages

(i) It Eliminates confusion while interpelling data

(ii) It Helps in deriving unnameable insights

(iii) It Offers data points to chart out a course of action.

Disadvantages

(i) It Can lead to a very large no.of tables

(ii) Lot of data on it that is not shown on the cross tab cross tabulation as a method of presenting and analysis of data.

(iii) Cross tabulation is a method to quantitatively analyze the relationship b/w variables.

(iv) Cross tabulation of categorical data can be done with through tools such as SPSS, Excel and SAS useful for studying market research.

(b) Explain how 'ogives' are drawn for any frequency distribution.

Ans :

The Ogive is defined as the frequency distribution graph of a series. The Ogive is a graph of a cumulative distribution, which explains data values on the horizontal plane axis and either the cumulative relative frequencies, the cumulative frequencies or cumulative per cent frequencies on the vertical axis.

Cumulative frequency is defined as the sum of all the previous frequencies up to the current point. To find the popularity of the given data or the likelihood of the data that fall within the certain frequency range, Ogive curve helps in finding those details accurately.

Create the Ogive by plotting the point corresponding to the cumulative frequency of each class interval. Most of the Statisticians use Ogive curve, to illustrate the data in the pictorial representation. It helps in estimating the number of observations which are less than or equal to the particular value.

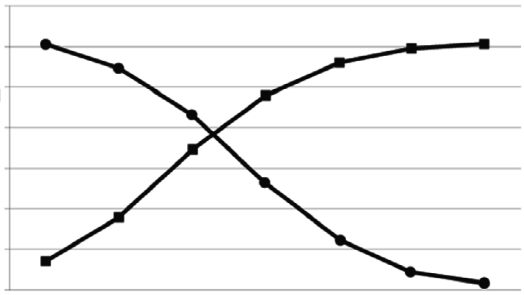
Ogive Graph

The graphs of the frequency distribution are frequency graphs that are used to exhibit the characteristics of discrete and continuous data. Such figures are more appealing to the eye than the tabulated data. It helps us to facilitate the comparative study of two or more frequency distributions. We can relate the shape and pattern of the two frequency distributions.

The two methods of Ogives are :

(i) Less than Ogive

(ii) Greater than or more than Ogive



The graph given above represents less than and the greater than Ogive curve. The rising curve represents the less than Ogive, and the falling curve represents the greater than Ogive.

(i) Less than Ogive

The frequencies of all preceding classes are added to the frequency of a class. This series is called the less than cumulative series. It is constructed by adding the first-class frequency to the second-class frequency and then to the third class frequency and so on. The downward cumulation results in the less than cumulative series.

(ii) Greater than or More than Ogive

The frequencies of the succeeding classes are added to the frequency of a class. This series is called the more than or greater than cumulative series. It is constructed by subtracting the first class, second class frequency from the total, third class frequency from that and so on. The upward cumulation result is greater than or more than the cumulative series.

5. (a) A stock market analyst wants to estimate the average return of a certain stock. A random sample of 15 days yields an annualized average return of × bar = 10.37% and a standard deviation of s = 3.5%. Assuming a normal population of returns, give a 95% confidence interval for the average return on this stock. **(Unit-III, Q.No.19)**

Sol :

The critical value of t for degree of freedom = (n – 1) = (15 – 1) = 14

Right tail area of 0.025 is = 2.145

The corresponding confidence interval estimate

$$= \overline{x} \pm t_{00.25} \frac{\sigma}{\sqrt{\eta}}$$

$$= 10.37 \pm 2.145 \frac{3.5}{\sqrt{15}}$$

$= 10.37 \pm 1.94$

$= 10.37 + 1.94 = 12.31$

$= 10.37 - 1.94 = 8.43$

(b) What is

(i) pie-chart **(Unit-III, Q.No.19)**

(ii) histogram ? **(Unit-III, Q.No.15 (1st Point))**

6. The following data pertain to the number of units of a product manufactured per day from four different brands of machines.

Machine Brands			
A	**B**	**C**	**D**
46	40	49	38
48	42	54	45
36	38	46	34
35	40	48	35
40	44	51	41

Test whether the mean productivity is the same for the four brands of machine type.

Sol :

Given No. of samples (k) = 4 (A, B, C, D)

Null hypothesis :

There is a no significant differences between mean Productivity of four brands of Machines.

Alternative hypothesis :

There is a significant differences between mean productivity of four brands of Machines.

Step 1 :

Calculation of variance between samples

X_1	X_2	X_3	X_4	Total
46	40	49	38	173
48	42	54	45	189
36	38	46	34	154
35	40	48	35	158
40	44	51	41	176
$\bar{X}_1 = 205$	$\bar{X}_2 = 204$	$\bar{X}_3 = 248$	$\bar{X}_4 = 193$	850

$$\text{Correction factor (CF)} = \frac{(\text{Grand Total})^2}{N} = \frac{(850)^2}{20} = \frac{722500}{20}$$

$$= 36{,}125$$

Total Sum of Square (TSS)

$$TSS = \sum_i \sum_j X_{ij}^2 - CF$$

$(46)^2 + (40)^2 + (49)^2 + (38)^2 + (48)^2 + (42)^2 + (54)^2 + (45)^2 + (36)^2 + (38)^2 + (46)^2 + (34)^2 + (35)^2 + (40)^2 + (48)^2 + (35)^2 + (40)^2 + (44)^2 + (51)^2 + (41)^2 + - 36{,}125$

$\Rightarrow$ $2116 + 1600 + 2401 + 1444 + 2304 + 1764 + 2916 + 2025 + 1296 + 1444 + 2116 + 1156 + 1225 + 1600 + 2304 + 1225 + 1600 + 1936 + 2601 + 1681 - 36{,}125$

$36{,}754 - 36{,}125 = 629$

Sum of Squares (SSC)

$$SSC = \sum_i \frac{T_j^2}{n_j} = \frac{(GT)^2}{N}$$

$$\left[\frac{(205)^2}{5} + \frac{(204)^4}{5} + \frac{(248)^2}{5} + \frac{(193)^2}{5}\right] - 36{,}125$$

$= 8405 + 8323.2 + 12300.8 + 7449.8 - 36125$

$= 36528.8 - 36125$

$= 403.8$

Sum of Squares between Rows

$$= \frac{(173)^2}{4} + \frac{(189)^2}{4} + \frac{(154)^2}{4} + \frac{(158)^2}{4} + \frac{(176)^2}{4} - CF$$

$\Rightarrow$ $7482.25 + 8930.25 + 5929 + 6241 + 7744 - 36125$

$= 36326.5 - 36125$

$= 201.5$

Residual error $= TSS - (SSC + SSR)$

$629 - 403.8 - 201.5 = 23.7$

Annova Table

Source of Variation	Sum of Squares	Degree of Freedom	Mean Variance	Variance Ratio
Sum of squares (between Column)	403.8	(c – 1) (4 – 1) = 3	$\frac{403.8}{3} = 134.6$	$\frac{134.6}{6.14} = 21.92$
Sum of square (between Rows)	201.5	(r – 1) = (5 – 1) = 4	$\frac{201.5}{4} = 50.375$	$\frac{50.375}{6.14} = 8.20$
Residual errors	23.7	(c – 1) (r – 1) 3 × 4 = 12	$\frac{23.7}{12} = 1.975$	—
	629	19		

$F_{0.05}$ (4.12) = 3.26 and $F_{0.05}$ (3.12) = 3.49

Conclusions :

1. $F_1 > F_{0.05}$ (4, 12). Hence H_0 is accepted that is the 5 workers differ respect to mean productivity.
2. $F_2 > F_{0.05}$ (3, 12). Hence H_0 is rejected that is mean productivity is not same.

7. (a) What care must be taken in designing capitalization in report preparation?

Ans :

- While determining the length of the report (since research reports vary greatly in length), one should keep in view the fact that it should be long enough to cover the subject but short enough to maintain interest. In fact, report-writing should not be a means to learning more and more about less and less.
- A research report should not, if this can be avoided, be dull; it should be such as to sustain reader's interest.
- Abstract terminology and technical jargon should be avoided in a research report. The report should be able to convey the matter as simply as possible. This, in other words, means that report should be written in an objective style in simple language, avoiding expressions such as "it seems," "there may be" and the like.
- Readers are often interested in acquiring a quick knowledge of the main findings and as such the report must provide a ready availability of the findings. For this purpose, charts, graphs and the statistical tables may be used for the various results in the main report in addition to the summary of important findings.
- The layout of the report should be well thought out and must be appropriate and in accordance with the objective of the research problem.
- The reports should be free from grammatical mistakes and must be prepared strictly in accordance with the techniques of composition of report-writing such as the use of quotations, footnotes, documentation, proper punctuation and use of abbreviations in footnotes and the like.

- The report must present the logical analysis of the subject matter. It must reflect a structure wherein the different pieces of analysis relating to the research problem fit well.
- A research report should show originality and should necessarily be an attempt to solve some intellectual problem. It must contribute to the solution of a problem and must add to the store of knowledge.
- Towards the end, the report must also state the policy implications relating to the problem under consideration. It is usually considered desirable if the report makes a forecast of the probable future of the subject concerned and indicates the kinds of research still needs to be done in that particular field.
- Appendices should be enlisted in respect of all the technical data in the report.
- Bibliography of sources consulted is a must for a good report and must necessarily be given.
- Index is also considered an essential part of a good report and as such must be prepared and appended at the end.
- Report must be attractive in appearance, neat and clean, whether typed or printed.
- Calculated confidence limits must be mentioned and the various constraints experienced in conducting the research study may also be stated in the report.

(b) What is 'scatter diagram'? **(Unit-IV, Q,No.14)**

(c) What are the components of 'time series'" **(Unit-V, Q,No.3)**

8. The following data are June 2018 to June 2019 commodity price index for a category of goods : 142, 137, 143, 142, 145, 151, 147, 144, 149, 154, 148, 153, 154. Form a new index, using January 2019 as the base month.

Sol :

Month	**Price Index**	**New price Index**
June - 2018	142	$\frac{142}{144}\times100$ = 98.61
July - 2018	137	$\frac{137}{144}\times100$ = 95.13
August - 2018	143	$\frac{143}{144}\times100$ = 98.30
September - 2018	142	$\frac{142}{144}\times100$ = 98.61
October - 2018	145	$\frac{145}{144}\times100$ = 100.69
November - 2018	151	$\frac{151}{144}\times100$ = 104.86

December - 2018	147	$\frac{147}{144}\times 100$ = 102.08
January - 2019	144 Base value	$\frac{144}{144}\times 100$ = 100
February - 2019	149	$\frac{149}{144}\times 100$ = 103.47
March - 2019	154	$\frac{154}{144}\times 100$ = 106.94
April - 2019	148	$\frac{148}{144}\times 100$ = 102.77
May - 2019	153	$\frac{153}{144}\times 100$ = 106.25
June - 2019	154	$\frac{154}{144}\times 100$ = 106.94

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

MBA I-Semester Examinations **R19**

January - 2020

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 3 Hours] [Max. Marks : 75

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

Part - A (5 × 5 Marks = 25)

ANSWERS

1. a) What are the features of a good research study? **(Unit-I, SQA - 2)**

b) List and describe some sources of primary data collection. **(Unit-II, Q.No. 14)**

c) Differentiate between univariate and multivariate data.

Ans :

Differences between univariate and bivariate data.

Univariate Data	Bivariate Data
➢ It involving a single variable	➢ It involving two variables
➢ It does not deal with causes orrelationships	➢ It deals with causes or relationships
➢ The major purpose of univariateanalysis is to describe	➢ The major purpose of bivariate analysis is to explain
➢ Central tendency - mean, mode, median	➢ Analysis of two variables simultaneously
➢ Dispersion - range, variance, max, min, quartiles, standard deviation.	➢ Correlations comparisons, relationships, causes explanations
➢ Bar graph, histogram, pie chart, line graph, box-and-whisker plot	➢ Tables where one variable is contingent on the values of the other variable.

d) When is a t-test used? What are its different types? **(Unit-III, Q.No. 21, 23, 25)**

e) What are some problems that are encountered while constructing index numbers? **(Unit-V, Q.No. 15)**

Part - B (5 × 10 Marks = 50)

2. What is meant by research? What are the objectives of research and its managerial value? What are the different types of research? Discuss in detail. **(Unit-I, Q.No. 1, 4)**

OR

3. Describe the research process in detail. Take an example of doing market research before launching a new product. **(Unit-I, Q.No. 5)**

4. What are the features of a good research design? **(Unit-II, Q.No. 10)**

OR

5. What is a research design? Discuss the different types of common research designs. **(Unit-II, Q.No. 3, 7)**

6. Define tabulation. Explain in delail the different parts of a table. **(Unit-III, Q.No. 1, 2)**

OR

7. What is a dependent Sample or repeated measures t-test? Explain its use by giving a suitable example. **(Unit-III, Q.No. 25)**

8. What is ANOVA? How is an ANOVA table setup? **(Unit-IV, Q.No. 1, 2, 3)**

OR

9. Use the sample data below to test the hypotheses
H_0 : p1 = p2 = p3
H_1 : Not all population proportions are equal

Where p_i is the population proportion of Yes response for population i. Using a 0.05 level of significance, what is the p-value and what is your conclusion?

Sol :

(i) Null Hypothesis (H_0)

All population proportions are equal

$$H_0 : P_1 = P_2 = P_3$$

(ii) Alternative Hypothesis (H_1)

Not all population proportions are equal

$$H_1 : P_1 \neq P_2 \neq P_3$$

(iii) Level of Signiface (α)

0.05 (1.96 Table value)

(iv) Test Statistics

$$Z_\alpha = \frac{P_1 - P_2 - P_3}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}\right]}}$$

(v) Conclusion :

In sufficient Data is given so that we can't conclude that the given proportions are equal (or) not

10. What do you understand by Exploratory Factor Analysis? Explain its use-case and utility for research by giving a suitable example. **(Unit-IV, Q.No. 36)**

OR

11. Consider the following time series data.

Week	1	2	3	4	5	6
Units	18	13	16	11	17	14

Develop a three-week moving average forecasts for this time series. Compute MSE and a forecast for week 7. Use alpha 0.2 to compute exponential smoothing forecasts for the time series.

Sol :

Week	Units	3 - Yearly Total	3 - Yearly M. Average
1	18	–	–
2	13	47	15.66
3	16	40	13.33
4	11	44	14.66
5	17	42	14
6	14	–	–

New Base = Previous Base + α (New Demand – Previous Base)

$$S_t = S_{t-1} + \alpha (D_t - S_{t-1})$$

$$S_7 = 17 + 0.2 \ (18 - 17)$$

$$17 + 0.2 (1) = 17.2$$