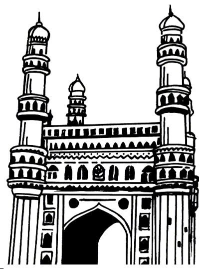
Rahul's ✔

Topper's Voice

R22
New Syllabus



J N T U (H)

M B A

I Year I Semester

Latest **2023** Edition

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

- **Study Manual**
- **Internal Assessment**
- **FAQ's and Important Questions**
- **Short Question & Answers**
- **Choose the Correct Answer**
- **Fill in the blanks**
- **Exercise Problems**
- **Solved Model Papers**
- **Solved Previous Question Papers**

Price
` . 249-00

- by -

WELL EXPERIENCED LECTURER

Rahul Publications ™

Hyderabad. Cell : 9391018098, 9505799122

J N T U (H) M B A

I Year I Semester

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Price `. 249 -00

Sole Distributors : **Cell : 9391018098, 9505799122**

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: Meaning, Scope, Role of Business Research, Types of Research, Research Process, Conceptualization of Variables and Measurement, Types and Measurement of Variables, 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, Characteristics of Good Research Design, Sampling and its Applications. Data Collection Methods & Tools: Types of Data, Sources and Instruments for Data, Guidelines for Questionnaire, Sampling and its Application. Measurement and Scaling, Reliability and Validity in Measurement of Variables, Sources of Error in Measurement.

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. Introduction to Statistics, Measurement of Central Tendency and Dispersion.

(b) **Small Sample Tests:** t-Distribution, Properties and Applications, Testing for One and Two Means, Paired t-Test, Hypothesis Formulation and Testing.

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.

UNIT - V

Time Series Analysis and Report Writing: Components, Models of Time Series, Additive, Multiplicative and Mixed Models, Trend Analysis: Free hand Curve, Semi Averages, Moving Averages, Least Square Methods.

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. 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

Frequently Asked & Important Questions

UNIT - I

1. What are the features of good research?

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

Refer Unit-I, Q.No. 2

---

2. What are the unique challenges in conduct of social science research ?

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

Refer Unit-I, Q.No. 9

---

3. What are the different types of research? Discuss in detail.

Ans : **(Jan-20)**

Refer Unit-I, Q.No. 11

---

4. Briefly explain the step-by-step process of research.

Ans : **(May-22, Jan.-20, Imp.)**

Refer Unit-I, Q.No. 13

---

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

Ans : **(Imp.)**

Refer Unit-I, Q.No. 16

---

6. What are the characteristics of good measurement tool?

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

Refer Unit-I, Q.No. 18

---

7. Define scaling. Explain the properties of scales.

Ans : **(Imp.)**

Refer Unit-I, Q.No. 21

---

8. What are the applications and dis- ad

Ans : **(May-22, Aug.-21, Nov.-20,Imp.)**

Refer Unit-I, Q.No. 22

---

9. What are the differences between Nominal, Ordinal, Interval and Ratio Scales ?

Ans : **(May-22)**

Refer Unit-I, Q.No. 23

10. What are the Ethical issues concerning research Participants ?

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

Refer Unit-I, Q.No. 24

UNIT - II

1. What are the sources of research problem ?

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

Refer Unit-II, Q.No. 3

2. Briefly explain Research Problems in the fields of marketing, production, finance and personnel management.

Ans : **(May-22)**

Refer Unit-II, Q.No. 4

3. Discuss the different types of common research design.

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

Refer Unit-II, Q.No. 7

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

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

Refer Unit-II, Q.No. 8

5. Explain "Before-after experimental design with control group".

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

Refer Unit-II, Q.No. 10

6. Briefly explain random and non-random sampling.

Ans : **(March-22)**

Refer Unit-II, Q.No. 15

7. What are the Applications of Sampling?

Ans : **(Imp.)**

Refer Unit-II, Q.No. 18

8. Explain the various methods of collecting primary data.

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

Refer Unit-II, Q.No. 21

9. What are the methods for collecting secondary data?

Ans : **(March-22, Nov.-21)**

Refer Unit-II, Q.No. 24

10. What are the Guidelines / Precautions for Questionnaire ?

Ans : **(March-22, Imp.)**

Refer Unit-II, Q.No. 31

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

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

Refer Unit-II, Q.No. 37

12.

Refer Unit-II, Q.No. 38

UNIT - III

1. Define tabulation. State the objectives of tabulation.

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

Refer Unit-III, Q.No. 4

2. Explain in detail the differnet parts of tables.

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

Refer Unit-III, Q.No. 5

3. What do you understand bivariate / multivariate tabulation.

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

Refer Unit-III, Q.No. 7

4. What is frequency distribution? Explain the formation of discrete and continuous frequency distribution.

Ans :

Refer Unit-III, Q.No. 9

5. For the given data below on the monthly income of 600 families of a rich locality in Hyderabad City, draw :

(i) Less than Ogive and

(ii) More than Ogive Curves and point out the Median.

Monthly Income (Rs. Crore)	No. of Families
Below 10	50
10 - 20	100
20 - 30	200
30 - 40	150
40 - 50	50
50 - 60	30
60 - Above	20

Sol : **(May-22)**

Refer Unit-III, Prob. 2

6. What do you mean by graphical repre-sentation of data ? State the rules for graphing. What are the advantages and disadvantages of graphs ?

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

Refer Unit-III, Q.No. 13

7. Explain the functions of statistics.

Ans : **(Imp.)**

Refer Unit-III, Q.No. 24

8. Define Small sample tests and t-distribution.

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

Refer Unit-III, Q.No. 64

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

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

Refer Unit-III, Q.No. 65

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

Refer Unit-III, Prob. 37

10. 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).

Sol : **(Aug.-21)**

Refer Unit-III, Prob. 38

10. What is dependent sample (or) repeated measures t-test.

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

Refer Unit-III, Q.No. 68

UNIT - IV

1. What is an ANOVA? State the assumptions and applications of an ANOVA.

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

Refer Unit-IV, Q.No. 1

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

Refer Unit-IV, Prob. 3

3. The following table gives the classification of 100 workers according sex and nature of work. Test whether the nature of work is independent of the sex of the worker.

	Stable	Unstable	Total
Male	40	20	60
Females	10	30	40
Total	50	50	100

Sol : **(Aug.-21)**

Refer Unit-IV, Prob. 9

4. What are the different types of correlations?

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 10

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

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 13

6. Explain briefly about Least Square Fit of Linear Regression.

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 23

7. Define multiple regression analysis. State its objectives and assumptions.

Ans : **(May-22)**

Refer Unit-IV, Q.No. 24

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

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 28

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

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

Refer Unit-IV, Q.No. 33

UNIT - V

1. Define Trend Analysis. Explain the purpose of measuring trend.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 6

2. Explain in detail semi average method with an example.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 10

3. Discuss the method of moving averages in measuring trend. What are it merits and limitations of moving average method?

Ans : **(Imp.)**

Refer Unit-V, Q.No. 11

4. What is least square method and explain its advantages and disadvantages?

Ans : **(Imp.)**

Refer Unit-V, Q.No. 12

5. Fit a straight line by the Least Square Method and tabulate the trend values for the above data.

Year	2011	2012	2013	2014	2015	2016	2017
Production (in tons)	77	88	94	85	91	98	90

Sol : **(June-19, Imp.)**

Refer Unit-V, Prob. 7

6. Explain the importance of index numbers.

Ans : **(May-22)**

Refer Unit-V, Q.No. 15

7. Describe briefly applications of index numbers in business and industry.

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

Refer Unit-V, Q.No. 16

8. What are the problems involved in construction of index numbers ? Explain.

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

Refer Unit-V, Q.No. 23

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

Ans : **(May-22, Imp.)**

Refer Unit-V, Q.No. 27

10. Explain the process of Report Preparation and Presentation.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 29

10. What is Documentation. Explain the Guides for Effective Documentation.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 34

10. What care must be taken in designing capitalization in report preparation?

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

Refer Unit-V, Q.No. 35

UNIT I

Introduction to Research: Meaning, Scope, Role of Business Research, Types of Research, Research Process, Conceptualization of Variables and Measurement, Types and Measurement of Variables, Ethics in Business Research.

1.1 Introduction to Research

1.1.1 Meaning

Q1. Define Research.

(OR)

What is meant by research ?

Ans : **(Jan-20)**

Introduction

Research is a scientific investigation. Investigation means a search for new facts and ideas in any branch of knowledge. Thus, we can say that research is a search for knowledge. Research may be considered as a movement, a movement from the unknown to the known. It is actually a voyage of discovery.

Research is carried out for two purposes; one is the discovery of new facts and the second, verification of the old ones. The object of every business organization, of course, is the discovery of new facts, new relationship, and new laws governing the business phenomena. But constant verification of the old concepts is also needed especially in dynamic business environment.

Meaning

In order to plan and carry out research, it is necessary to know what we mean by research-in general, as well as in the specialized fields of business management.

"Research is an Organized and Systematic way of Finding answers to Questions."

Systematic because there is a definite set of procedures and steps which you will follow.

There are certain things in the research process that are always done in order to get the most accurate results.

Organized in that there is a structure or method in going about doing research. It is a planned procedure, not a spontaneous one. It is focused and limited to a specific scope.

Finding answers is the end of all research. Whether it is the answer to a hypothesis or even a simple question, research is successful when we find answers. Sometimes the answer is no, but it is still an answer.

Questions are central to research. If there is no question, then the answer is of no use. Research is focused on relevant, useful, and important questions. Without a question, research has no focus, drive, or purpose.

The word research is derived from the Latin word meaning to know. It is a systematic and a replicable process, which identifies and defines problems, within specified boundaries. It employs well-designed method to collect the data and analyses the results. It disseminates the findings to contribute to generalizable knowledge.

Definitions

Various social and behavioural scientists have defined the word research in different ways. Some of the most popular definitions are:

(i) According to Pocket Oxford Dictionary

"Endeavour to discover facts by scientific study, course of critical investigation".

(ii) According Collins Concise Dictionary

"Systematic investigation to establish facts or collect information on the subject".

(iii) According to Redman and Mory "Research is systematized effort to gain new knowledge".

(iv) According to Encyclopedia of Social Sciences"Research is the manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art".

Thus, we can say that research is a systematic and objective attempt to study a business problem for the purpose of deriving general principles. In other words, research is a systematic, controlled, empirical and critical investigation of hypothetical propositions about the presumed relations among phenomena.

Q2. Enlist various characteristics of research.

(OR)

What are the features of good research?

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

Characteristics

A well-organized research must possess certain characteristics features, which are as follows:

1. Solution Oriented

The problem of research must be clearly defined and stated. The motive of research must be mentioned in the beginning of research work. The research should provide a solution of a business problem.

2. Logical

In a research we find out facts about a phenomenon and draw conclusions about it. The inferences and generalizations thus made must be logical. For example, all illiterate people in the village live longer than the educated people in the cities leads to the conclusions that illiteracy is the cause of longevity. This is an example of illogical research conclusion.

3. Objective

Observing true picture of a phenomenon without being affected by observers own opinion is termed as 'objective'. Objectivity means knowing reality. The criterion of objectivity is that all researchers should arrive at the same conclusion about the phenomenon on which they are pursuing research.

4. Impartiality

A dishonest research may select data items of individuals to draw conclusions to his favour. This brings bias into research, which affects the objective of the study. Therefore a true research must be impartial and unbiased.

5. Accuracy

A research worker needs to gain some expertise in the study he is undertaking. This expertise results in achieving the accuracy in the solution drawn. The accuracy of conclusions is a sensitive issue as it may affect the whole decision-making.

6. Systematic

In a research there should be well-defined steps. Each step should be sequentially linked with another, so that, the whole research work is an organized structure.

7. Verifiability

the results of a research are subjective to verifications. For building a sound basis for decision making one verifies the research results by replicating the study.

8. Empirical

A research is an empirical process and involves data collection. The results are based on observed experience or empirical evidence. Research rejects assumptions and dogma as methods of established knowledge. It accepts only what is verified by empirical observations.

Q3. "Many a time management is not convinced about the utility of research and regards it as an unnecessary activity over which no funds should be spent". Comment on this statement explaining the objectives of research.

(OR)

What are the Objectives of research?

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

Objectives

The main goal of research is to improve the quality and level of living in the society. The purpose of a research study is to find out the hidden facts about a business phenomenon. The obvious function of research is to add new knowledge to the existing store. It serves the government and the business organizations in forming their future policies.

The objectives of a research study are listed below:

1. Understanding a business problem

The first and foremost objective of any study is to understand, analyze and explore a business problem. Once complete familiarity with the phenomenon is achieved, it is easier to decompose the complex problem into smaller once.

2. Identifying the cause and effect relationship

Individuals form groups, and groups form organizations. They are interdependent. It is very important for a researcher to identify the functional relationships among various components of an organization. A scientific investigation is necessary in studying the cause and effect relationship of variables involved in a business phenomenon.

3. To innovate new ideas

One of the objectives of a researcher is to bring constant improvement in the techniques of his trade. Apart from verifying and testifying the existing assumptions, one of the functions of a research is to add new knowledge to the state of the art. Research invokes the innovation of new concepts, theories and idea in a business study. Apart from this, research also removes and discards worthless theories that are prevalent in the society.

4. To improve the quality

The whole exercise of any activity is done for the improvement of quality of a product, machinery, or life of human beings. For a business organization it is almost important to improve the quality of its products. This can be achieved by a systematic and critical investigation i.e. research.

Q4. Who are the stakeholders in research ?

(OR)

Who are the participants in research ?

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

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.

1.1.2 Scope

Q5. 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 analyzed 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 specialization 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 maximizing of profit, and production functions have to minimize 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 organizational 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.

Q6. What is Social Science Research? Explain its nature.

Ans :

Definitions

(i) According to Prof. Ruimnel "Social Science research is devoted to a study of mankind in his social environment and is concerned with improving his understanding of social orders, groups, instructions and ethics."

(ii) According to Prof. M.H. Gopal "Social Science is the scientific analysis of the nature and trends of social phenomenon or groups or in general of human behaviour so as to formulate broad principles and scientific concepts."

Nature

1. Social science research deals with the social phenomenon. It studies the behaviour of human beings. It covers the study of economic, political, educational, administrative and related aspects of social life.
2. Social science research aims at discovery of new facts and verifications of old ones.
3. Social science research tries to establish causal relationship between various human activities. It discovers the rules or laws so that they can be used in the guided growth of human society.
4. Social science research involves the use of scientific methodology. It implies the development of new scientific tools, concepts, theories which would facilitate reliable and valid study of human behaviour.
5. Social science research assists in the understanding of evolution of new theories. Every research highlights some broad principles, establishes some scientific truth, analyses their sequences.
6. In social science research, human beings are the sources of data. So. their attitudes, environments, honesty, personal values and biases can change the research decisions.
7. Social science research is utilitarian in nature. Here, the primary goal is to understand social life and thereby, gain greater knowledge which will be helpful for his project or to control over social behaviour.
8. The purpose of social science research can be academic. It means to acquire first hand knowledge about the society. To know and understand the laws, to intimate knowledge of human society are the purposes of social science research.

Q7. Explain the significance of social science research?

Ans :

Utility

1. Social Control

With the help of social science research, one can go for planning. A control over society is possible only when one has complete knowledge of the organization and working of society and its various institutions. All these can be achieved only through a scientific study of society.

2. Social Cohesion

The study of society creates better under standing and social cohesion between different groups. It reveals the underlying unity in the midst of disparity and creates a feeling of oneness.

3. Social Welfare

Social welfare can be achieved through social research. Such research helps us to judge the magnitude of social evils and thus, take necessary steps to remove them.

4. Social Prediction

Social science research helps the society in making predictions and goals for the future. It becomes possible because they have necessary data at their disposal, Statistical methods can be used for prediction.

5. Social Growth

Social science research aims at opening new visitors of growth; knowledge, wisdom. The future path of social progress is conditioned by our knowledge of ourselves and other people. Social science research helps in guiding the trend of social growth on proper lines.

Q8. Explain the Need of Social Science Research ?

Ans :

Social science research needs interdisciplinary approach in handling problems or situations. There is a need to blend various social sciences to deal with problems, i.e.. social problems in a particular society. The interdisciplinary approach facilitates better understanding and management of the complex problems or situations facing a particular society.

The need for multi disciplinary and interdisciplinary approach in social science research can be explained below:

1. Limitations of Individual Social Science

There are limitations of individual social sciences. Sociology as a discipline would alone not help to solve the problem of poverty. The problem of poverty has to be seen from the angle of economics, politics, etc.

2. Interrelationships among social sciences

There is a need for interrelationships among social sciences. For instance, to solve the problem of unemployment in rural areas, the Govt, may adopt certain employment schemes or projects. Such decision can be taken from economic point of view or social point of view or even from politics point of view.

Thus, there should be proper mix of two related social sciences.

3. Need for objectivity in research findings

There are certain problems or issues in social sciences, which can be dealt only through multi disciplinary or interdisciplinary approach. The problem must be studied with discipline. Then only it would affect the objectivity of conclusions drawn from such study.

Thus, overall to solve complexity of social phenomenon and for getting objectivity in research findings, there is a need to adopt interdisciplinary as well as multi disciplinary approach in research.

Q9. What are the unique challenges in conduct of social science research ?

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

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.

1.1.3 Role of Business Research

Q10. Define Business Research. Explain the role of Business Research.

Ans :

Meaning

Business research may be defined as research activities carried out relating to the different functionaries in the business and corporate world. Business research is a process of acquiring detailed information of all the areas of business and using such information in maximizing the sales and profit of the business. Such a study helps companies determine which product/service is most profitable or in demand. The definition of business research involves acquiring information and knowledge for professional or commercial purposes such as determining opportunities and goals for a business. An example of business research is gathering sales information and writing a detailed report on marketing and sales.

Business research is a systematic inquiry that provides information to guide managerial decisions. In other words, it is a process of planning, acquiring, analyzing, and disseminating relevant data, information, and insights to decision makers in ways that mobilize the organization to take appropriate actions that, in turn, maximize performance.

Role

1. Purpose Clearly Defined

The purpose of the business research the problem involved or the decision to be made should be clearly defined and sharply delineated in terms as unambiguous as possible. The statement of the decision problem should include its scope, its limitations, and the precise meanings of all words and terms significant to the research. Failure of the researcher to do this adequately may raise legitimate doubts in the minds of research report readers as to whether the researcher has sufficient understanding of the problem to make a sound proposal attacking it.

2. Research Process Detailed

The research procedures used should be described in sufficient detail to permit another researcher to repeat the research. This includes the steps to acquire participants, informed consent, sampling methods and representativeness, and data gathering procedures. Omission of significant procedural details makes it difficult or impossible to estimate the validity and reliability of the data and justifiably weakens the confidence of the reader in the research itself as well as any recommendations based on the research.

3. Research Design thoroughly Planned

The procedural design of the research, and its choice among competing designs, should be clearly described and carefully planned to yield results that are as objective as possible. A survey of opinions or recollections ought not to be used when more reliable evidence is available from documentary sources or by direct observation. Bibliographic searches should be as thorough and complete as possible. Experiments should have satisfactory controls, reducing threats to internal validity and enhancing the probability of external validity (generalizability). Direct observations should be recorded as soon as possible after the event. Efforts should be made to minimize the influence of personal bias in selecting and recording data.

4. High Ethical Standards Applied

Researchers often work independently and have significant latitude in designing and executing projects. A research design that includes safeguards against causing mental or physical harm to participants and makes data integrity a first priority should be highly valued. Ethical issues in research reflect important moral concerns about the practice of responsible behavior in society. Careful consideration must be given to those research situations in which there is a possibility for physical or psychological harm, exploitation, invasion of privacy, and/or loss of dignity. The research need must be weighed against the potential for these adverse effects. Typically, you can redesign a study, but sometimes you cannot. The researcher should be prepared for this dilemma.

5. Limitations Frankly Revealed

The researcher should report, with complete frankness, flaws in procedural design and estimate their effect on the findings. There are very few perfect research designs. Some of the imperfections may have little effect on the validity and reliability of the data; others may invalidate them entirely. A competent researcher should be sensitive to the effects of imperfect design. The researcher's experience in analyzing data should provide a basis for estimating the influence of design flaws. As a decision maker, you should question the value of research about which no limitations are reported.

1.2 Types of Research

Q11. What are the different types of research? Discuss in detail.

Ans : **(Jan-20)**

Types

Research is a multidimensional activity. It comes in various forms and is used in all social, behavioral, educational, economical and management sciences.

According to the approach and method involved in a research, one can classify the following types of research.

1. Descriptive v/s Analytical research

Descriptive research basically describes what is. It mainly involves collection, recording, describing and analyzing the facts related to the study. It tries to find the existing status, trend and state of affairs in a phenomenon. Descriptive research involves surveys, but they are not merely data collection as they also involve measurement, classification, analysis, comparison and interpretation. In this type of research the variable under study are uncontrollable. One can only observe and report what is happening in a situation.

Analytical research, on the other hand deals with what will be. In this type of research, the variables involved are carefully and scientifically controlled and manipulated. Analytical research is also known as experimental research and is a very sophisticated technique. This kind of research is based on four important characteristics namely; control, manipulation, observation and replication.

2. Applied v/s Fundamental research

Applied research is action oriented or solution oriented. The main goal of an applied research is to obtain an immediate, specific and practical solution of a problem that a business organization is facing right now. It gives here and now solutions in actual problem situations. It involves scientific investigations but the methods are not so rigorous as in fundamental research. It finds solutions to be applied in local environment and they may not be universally acceptable. Applied research does not promise to add new knowledge to the discipline.

Fundamental research is carried out to scientifically enhance the organized body of knowledge of a discipline. Also know as basic research, it is concerned with formulation of theory and generalizations of principles. To evaluate and expand a formulated theory it may use empirical data. Basic research involves systematic, highly sophisticated scientific techniques. Fundamental research may not suggest the solutions of immediate problems, it rather draws long term conclusions.

3. Quantitative v/s Qualitative research

Quantitative research is based on quantitative variables, which can be measured in appropriate units. These involve objects and individuals that vary in size, quantity, amount, scale or degree. For example, prices of commodity can be measured in rupees, weight of a product is measured in kilograms and the mileage of vehicle is measured in kilometers per liter.

Qualitative research, on the other hand, is based on qualitative variables, which vary in quality of type. These variables cannot be measured on a scale or in any units. Social scientists use qualitative research for studying human behaviour. In market research surveys qualitative research is carried out to investigate the likes and dislikes of customers. It helps in understanding the current pattern of demand of a company's products.

4. Conceptual v/s Empirical research

Conceptual research involves the development of new theories, abstract ideas, and generalized principles. Philosophers, intellectuals and thinkers carry out this kind of research. On the basis of their conceptual knowledge they build theoretical models.

Conceptual research is an intellectual process to develop and verify knowledge.

Empirical research is based on observation and experimentation. The information collected in the form of facts develops the conclusions and theories about a phenomenon.

The models, so developed, can again be verified by a replication of data collection. To test a given hypothesis empirical research is most popular and powerful tool in the modern world.

5. Other types of research

Any research study is derivation of one or the other of above four types of research. One can further classify a research on the basis of its purpose, time taken and the discipline of knowledge it relates to. For example, Historical research is the study of past events, historical documents, remains and relics. Clinical research is employed to study the effects of a new drug. Market research is performed to forecast the potential demand of a product. One- time research is carried out on a small scale in short period with a specific purpose. Educational research is directed towards the study and development of educational system. Social research is concerned with the social problems of the society. Field research is done by going out in the field or market, where as Laboratory research is carried out with in four walls of a laboratory.

Q12. What is meant by exploratory research? What are the purposes for which it is conducted?

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

Meaning

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

1.3 Researchs Process

Q13. Briefly explain the step-by-step process of research.

(OR)

Describe the research Process in detail.

Ans : **(May-22, Jan.-20, Imp.)**

Research is a search for knowledge. It helps in taking appropriate decisions. Research involves asking a question and then trying to find an answer to it. Research is essentially a systematic, scientific and structured inquiry seeking facts through objective methods. Therefore a research must have a clearly defined step-by-step process. A knowledge of the research process is essential both for those who conduct the research and for those who wish to be benefited by the conclusions drawn from the research. A meaningful knowledge should have a definite purpose and direction.

In developing a research process, one would like to list the sequence of step-by step activities. In a research process these steps are inter- dependent and may overlap each other. They may not follow a strict sequence and the researcher has to be vigilant of their order continuously through out the research process. However, one can broadly enlist the main steps involved in a research process as a procedural guideline to the researcher.

These steps are:

1. Problem formulation
2. Literature survey
3. Development of hypothesis
4. Research design
5. Choice of sample design
6. Data collection
7. Analysis and interpretation of data
8. Hypothesis testing
9. Interpretation of results
10. Report writing

The above procedure can be depicted in a diagrammatic form as shown in the flowchart in figure. A brief description of the above steps is given below.

1. **Problem formulation**

   Formulation of a problem is the first and foremost step in a research process. It is not always easy to identify and define a problem in an ever-changing business environment. A researcher not only discovers and defines a problem area but also a specific problem within that area concerning his interest in business. The problem should be clearly and precisely stated. The statement of the problem must be complete.

   The problems in a business may sometimes be obvious and one can pinpoint them. Many a problems is not so apparent and needs explorations. Thus, first of all one has to identify a problem specifically and thoroughly, and then it has to be expressed in scientific terms so that statistical analysis can be performed on that problem.

2. **Literature survey**

   After the formulation and identification of a problem, the next important step is the review of literature survey. An exhaustive and critical review of professional literature familiarizes the researcher with the current state of knowledge. It helps in understanding of the problems and hypothesis that others have studied. It clarifies the concepts, theories, major variables involved, operational definitions and research methods used in the past. This contributes to the cumulative nature of scientific knowledge.

   Every year thousands of articles, books and monographs are published in any field of study. Therefore, it is important to sort out the relevant literature connected with the field of one's interest. It is best to begin any search for literature with one of the guides to published literature. These guides are increasingly computerized and include bibliographies, indexes and abstracts.

   With the advent of Internet the modern life has changed drastically. One can find an ocean of information within the four walls of one's study room through Internet. Some of the popular search engines like google, yahoo and rediff are becoming more popular in searching for literature on any topic.

The Research Process Flowchart

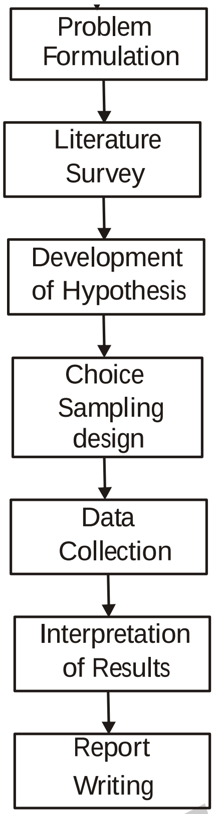


Fig. Research Process

3. Development of hypothesis

Once a problem is defined and a review of literature is made, the next step is to define clearly the hypothesis in a research problem. A hypothesis is a tentative assumption in a research problem, which has to be tested empirically with the help of observed data. When formulating a hypothesis, a researcher does not know whether it will be rejected or accepted. A hypothesis is constructed and tested; if it is rejected, another one is formulated; if it is accepted, it is incorporated in the scientific body of knowledge.

One should arrive at a clear and specific working hypothesis for which research methods already exist. A good hypothesis states a research problem in concise and precise terms so that the researcher is focused on the problem at hand.

4. Research design

A research design is a well-defined plan of action. It is a planned sequence of the entire research process. It is a blue print of research activity. In a big business scenario designing a research study is very complex. Therefore, a research design may change during the operation of a project. A good research design must use minimum of resources like time, money and manpower. A research design must be able to translate the general scientific model into a practical research operation. A scientifically developed research design possess the characteristics like

(i) Objectivity

(ii) Reliability

(iii) Validity

(iv) Generalization.

5. Choice of sample design

In any investigation the group of all items, objects or individuals under study is called ‘population’ or ‘universe’. For all purpose of determining population characteristics, instead of enumerating entire population, some items of the population, called a sample, are observed. Then the sample characteristics are utilized to approximately determine or estimate the population. For example, on examining the sample of particular product, we arrive at a decision of purchasing or rejecting that product. There will be surely some error in such an approximation and is inherent and unavoidable in any and every sampling scheme. But samples results in considerable gains in terms of time, money, accuracy and efforts.

Drawing a sample of some predeter- mined size from an entire population is not a child’s play. These have to be a systematic plan to choose the sample items. This plan or a technique of drawing a sample is known as sample design or sample plan or sampling technique. Researchers have suggested various sample designs. One research situation may be different from another, therefore, simple random sample, though most popular, may not be suitable in each case. Depending on the requirement of a situation one can choose one of the following sample designs:

1. Purposive or Judgment Sampling
2. Simple Random Sampling
3. Stratified Random Sampling
4. Systematic Sampling
5. Cluster Sampling
6. Area Sampling
7. Multi-stage Sampling
8. Multi-phase Sampling

6. Data collection

Once a sample design is formulated, the next step in a research process is the collection of relevant data. There can be two sources of data

(i) Internal data, that refers to the happenings and functions of a business organization. For example, the sale purchases details of a company.

(ii) External data, which is related to outside sources and external agencies. There are two types of data

(a) Primary data

(b) Secondary data.

(a) Primary data

Primary data are those observations, which are collected by an investigator for the first time.

(b) Secondary

Secondary data are already available in the records and have been collected by some other researcher for the purpose of studying a similar problem.

Methods of collecting primary data are:

(i) by direct observation and experimentations

(ii) by direct personal interview

(iii) by direct interviews through phone, sms and email.

(iv) by indirect personal interview

(v) by mailed / emailed Questionnaire

(vi) by schedules through enumerators

Methods of collecting secondary data:

(i) International organizations like WHO, UNO etc.

(ii) Government publications like economic survey, CSO, NSSO.

(iii) Journal and Newspapers

(iv) Research articles

(v) Reports of business organization and financial institutions.

A method and source of data collection is chosen by an investigator taking into account the objectives and requirements of the inquiry. The adopted method should incur minimum cost and time should have a reasonable level of accuracy and unbiased ness.

7. Analysis and interpretation of data

After the collection of data, what we have is a huge chunk of observations and numerical values. The data at the beginning are in raw form. For the purpose of applying further statistical techniques, one has to put the raw data in a useful form by classification, tabulation and categorization of data. If one has to feed the data in a computer, the data should bear the same form as required by the software used. This kind of processing of data involves one or more of the following activities:

(i) coding

(ii) labeling

(iii) editing

(iv) tabulation

(v) classification.

8. Hypothesis testing

After analyzing and processing of data, it is time now to test the hypothesis that were formed in step 3 of the research process. A hypothesis is skeptically formulated regarding the relationship between phenomena and variables involved in a study. Then by empirical investigation the hypothesis is tested for possible acceptance or rejection. In other words, the researcher decides on the basis of the observed facts that he has collected, whether or not an assumption is valid. A hypothesis is tested by making use of a predefined decision rules established in statistical methods. Some of the popular statistical tests are, Z- test, Chi- square test, t-test and F- test.

In a situation where no hypothesis is formulated in a study, the observations are made on the data directly and conclusions are drawn to formulate new generalizations and assumptions for future purposes.

9. Interpretations of results

After the data collection and testing of hypothesis one has to reach to the conclusions of the research study. These conclusions are the most vital outcomes of the study and have to be dealt with very carefully. On the basis of findings of the research work done we draw inferences about the phenomenon under study. This is a useful activity as without any outcome a research study is fruitless. The results obtained from the analysis of data are to be interpreted skillfully. A wrong interpretation my lead to wrong decisions. Interpretation may also lead to generalizations of the phenomena understudy. It may also help in developing new theories and can suggest new research problems to be explored in future.

10. Report writing

Last but not the least is the step of reporting the facts and findings of the research study. A report is a summary of the whole research process. The layout of a report must be attractive. The words used in the text must be easily comprehensive to a reader. Even a non- technical person understands a good report. In the beginning of the report one should give the title, time period of work, acknowledgement and preface. In the main text an introduction to the problem, summary of findings, results and inferences, and then the recommendations of the researcher are given. The report should conclude with appendices, bibliography and a subject or / and author indexes.

Q14. Take an Example of doing market research before Launching a new product.

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

Phase 1: New Product Idea Generation (Exploratory Research)

Exploratory research is a good starting point for any organization. One of the most important steps in new product development, it provides value, even if an organization is in the initial phases of concept testing.

This type of research gives an understanding of current customer satisfaction and provides direction for optimizing and marketing current product/service offerings.

However, the main goal of this research is to understand the general needs of the market and identify new product or service opportunities for your organization.

Phase 2: New Product Concept or Prototype Testing Research

After conducting exploratory research (formally or informally), an organization has identified a need within the market and is set to develop a new product that customers will be interested in.

The main goal of new product testing research is to have your target audiences evaluate the new product and ensure that the product meets their needs

Depending on the complexity of your product, this type of research may need to be re-visited at multiple points throughout the product development process.

In the beginning…

Early in the process, you may have a basic concept and idea that you want to test before proceeding. Conduct new product research now!

In the middle...

In a few months, your brand may have more detailed product specifications with renderings or product mock-ups. Again, conduct market research!

In the end...

Finally, near product design finalization, you may have an actual prototype or final product that the consumer can evaluate and provide feedback on. You guessed it – conduct prototype testing research!

Q15. "Creative management, whether in public administration or private industry, depends on methods of inquiry that maintain objectivity, clarity, accuracy and consistency". Discuss this statement and examine the significance of research.

Ans :

Significance

In any scientific inquiry research has been the most powerful tool for knowledge seeking people. Research is a kind of power with which one can foresee the implications of a particular phenomenon. Research is all pervading and is used in every discipline of study. Some of the applications of research are listed below.

1. In psychology, research is done to study psychological, cultural and motivational factors of different types of persons.
2. In sociology, research studies are performed to analyze changing relationships among age and sex groups, emerging class patterns, social mobility and social values etc.
3. In economics, planned development of a country and an organization is achieved through research. It helps us in deciding the wages, salaries of all employees, profit and risk involved in a business activity, and the effects of government policies on the economic structure.
4. In geography, research is used to study environmental control, climatic complexities, geographical appraisal of a country's foreign policy, geographical patterns of changes in agricultural wastelands.
5. In education, research helps in the improvement of teaching methods, interrelationship of teachers and students, expenditure share of primary, technical and higher education.
6. In medical science, clinical trials are performed to study the effects of medicine. No medicine or medical treatment is accepted by medical associations without completing a thorough research process on the medicine or treatment.
7. In the field of business management, marketing research is the backbone of marketing a product. Before launching a new product, market surveys are conducted to identify the needs and satisfaction level of customers. It helps not only in solving existing problems but also in identifying new opportunities. In every manufacturing industry there is a research and development section, which is involved in improving the quality of products. Research is carried out in production units to decide the amount, time and potential customers of its products. Research methods are applied in statistical quality control and maintaining the optimum inventory level. For a human resource manager it is important to have the knowledge of its employee's, salary structure, and satisfaction level, cost of living performance appraisal. The HR department achieves the objective through research studies
8. The government cannot function without having clear and true picture of what is happening in its state with its subjects. Every government department has a research officer with separate section on research. The collection of data is done round the year in every department and the government policies, budget, development programs take shape on the basis of these research surveys.

1.4 Conceptualization of Variables and Measurement

1.4.1 Types and Measurement of Variables

Q16. Define the term variable. Explain various types of variables.

Ans : **(Imp.)**

Meaning

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

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 independent 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.

Q17. Compare and contrast between independent variable and dependent variable.

Ans :

The differences between independent and dependent variables are follows,

S.No.	Independent variable	S.No.	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, regress and, measured, explained, experimental output variable.

Q18. Define measurement. What are the characteristics of measurement?

(OR)

What are the characteristics of good measurement tool?

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

Meaning

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.

Definitions

(i) **According to stevens** measurement is the assignment of numerals to objects or events according to rules.

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

Characteristics

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 on a test in a manner that allows everybody an equal chance to demonstrate their knowledge or skills.

Q19. 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.

Q20. 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

$$E = \text{errors}$$

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.

Q21. Define scaling. Explain the properties of scales.

Ans : **(Imp.)**

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 order 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 objects 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.

Q22. What are the different types of measurement scales with examples ?

(OR)

What are the applications and disadvantages of nominal, ordinal, interval and ratio scales ?

(OR)

What are the levels of measurements?

Ans : **(May-22, Aug.-21, Nov.-20,Imp.)**

(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.

(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.

(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.

(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."

Q23. Differentiate among nominal, ordinal, interval and ratio scales on any five points.

(OR)

Compare and contrast Nominal, Ordinal, Interval and Ratio Scales.

(OR)

What are the differences between Nominal, Ordinal, Interval and Ratio Scales ?

Ans : **(May-22)**

S.No.	Nominal Scale	Ordinal Scale	Interval Scale	Ratio Scale
1.	Nominal scale involves labels. It is a qualitative scale in which numbers are regarded as labels (or) tags for identifying and classifying the objects.	Ordinal scale is a ranking scale in which the numbers assigned to the objects represents its characteristics or positions.	Interval scale helps in comparing the differences taking place between the objects. The zero point and the units of measurement are taken randomly	Ratio scale helps the researchers to identify and rank the objects and compare the intervals (or) differences taking place between them The ratios of scale values are also computed with the help of this scale. This scale has 2 fixed zero point.
2.	Nominal scales are used for numbering the football players and social security.	Ordinal scales are used for ranking the teams in tournment, socio-economic class occupational status and quality rankings.	Interval scales are used for measuring the space temperature.	Ratio scales are used for measuring the height, weight, age and money of the individuals.
3.	In marketing research, nominal scales help in identifying the respondents, brands, attitudes, stores and other objects.	In marketing research, ordinal scales help in measuring the attitudes, opinions, perceptions and preferences of the individuals.	In marketing research, interval scales help in measuring the attitudes, opinions and index numbers.	In marketing research, ratio scales help in measuring the sales, costs, market share and number of variables.
4.	Nominal scales can be expressed as percentages and modes.	Ordinal scales can be expressed as percentile, quartile and median.	Interval scales can be expressed as range, median and standard deviation.	Ratio scales can be expressed as geometric mean and harmonic mean.
5.	The accuracy of nominal scales can be measured with the help of chi-square and binominal test.	The accuracy of ordinal scales can be measured with the help of rank order correlation, Friedman and ANOVA.	The accuracy of interval scales can be measured with the help of product moment correlations, t-tests ANOVA and regression factor analysis.	The accuracy of ratio scale: can be measured with the he of coefficient of variation.) :

1.5 Ethics in Business Research

Q24. Discuss the various ethical issues in business research.

(OR)

What are the Ethical issues concerning research Participants ?

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.

Q25. 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 convenient 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.

Q26. What is a diagnostic study ? What is the purpose of it ?

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

- 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.

Short Question & Answers

1. Define Research.

Ans :

Introduction

Research is a scientific investigation. Investigation means a search for new facts and ideas in any branch of knowledge. Thus, we can say that research is a search for knowledge. Research may be considered as a movement, a movement from the unknown to the known. It is actually a voyage of discovery.

(i) According to Pocket Oxford Dictionary

"Endeavour to discover facts by scientific study, course of critical investigation".

(ii) According Collins Concise Dictionary

"Systematic investigation to establish facts or collect information on the subject".

2. Features of good research.

Ans :

A well-organized research must possess certain characteristics features, which are as follows:

(i) Solution Oriented

The problem of research must be clearly defined and stated. The motive of research must be mentioned in the beginning of research work. The research should provide a solution of a business problem.

(ii) Logical

In a research we find out facts about a phenomenon and draw conclusions about it. The inferences and generalizations thus made must be logical. For example, all illiterate people in the village live longer than the educated people in the cities leads to the conclusions that illiteracy is the cause of longevity. This is an example of illogical research conclusion.

(iii) Objective

Observing true picture of a phenomenon without being affected by observers own opinion is termed as 'objective'. Objectivity means knowing reality. The criterion of objectivity is that all researchers should arrive at the same conclusion about the phenomenon on which they are pursuing research.

(iv) Impartiality

A dishonest research may select data items of individuals to draw conclusions to his favour. This brings bias into research, which affects the objective of the study. Therefore a true research must be impartial and unbiased.

(v) Accuracy

A research worker needs to gain some expertise in the study he is undertaking. This expertise results in achieving the accuracy in the solution drawn. The accuracy of conclusions is a sensitive issue as it may affect the whole decision-making.

3. 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.

4. Social Science Research? Explain its nature.

Ans :

Definitions

(i) According to Prof. Ruimnel "Social Science research is devoted to a study of mankind in his social environment and is concerned with improving his understanding of social orders, groups, instructions and ethics."

(ii) According to Prof. M.H. Gopal "Social Science is the scientific analysis of the nature and trends of social phenomenon or groups or in general of human behaviour so as to formulate broad principles and scientific concepts."

5. Need of Social Science Research.

Ans :

1. Limitations of Individual Social Science

There are limitations of individual social sciences. Sociology as a discipline would alone not help to solve the problem of poverty. The problem of poverty has to be seen from the angle of economics, politics, etc.

2. Interrelationships among social sciences

There is a need for interrelationships among social sciences. For instance, to solve the problem of unemployment in rural areas, the Govt, may adopt certain employment schemes or projects. Such decision can be taken from economic point of view or social point of view or even from politics point of view.

Thus, there should be proper mix of two related social sciences.

3. Need for objectivity in research findings

There are certain problems or issues in social sciences, which can be dealt only through multi disciplinary or interdisciplinary approach. The problem must be studied with discipline. Then only it would affect the objectivity of conclusions drawn from such study.

Thus, overall to solve complexity of social phenomenon and for getting objectivity in research findings, there is a need to adopt interdisciplinary as well as multi disciplinary approach in research.

6. Define Business Research.

Ans :

Meaning

Business research may be defined as research activities carried out relating to the different functionaries in the business and corporate world. Business research is a process of acquiring detailed information of all the areas of business and using such information in maximizing the sales and profit of the business. Such a study helps companies determine which product/service is most profitable or in demand. The definition of business research involves acquiring information and knowledge for professional or commercial purposes such as determining opportunities and goals for a business. An example of business research is gathering sales information and writing a detailed report on marketing and sales.

Business research is a systematic inquiry that provides information to guide managerial decisions. In other words, it is a process of planning, acquiring, analyzing, and disseminating relevant data, information, and insights to decision makers in ways that mobilize the organization to take appropriate actions that, in turn, maximize performance.

7. Analytical Research.

Ans :

Analytical research, on the other hand deals with what will be. In this type of research, the variables involved are carefully and scientifically controlled and manipulated. Analytical research is also known as experimental research and is a very sophisticated technique. This kind of research is based on four important characteristics namely; control, manipulation, observation and replication.

8. Fundamental Research.

Ans :

Fundamental research is carried out to scientifically enhance the organized body of knowledge of a discipline. Also know as basic research, it is concerned with formulation of theory and generalizations of principles. To evaluate and expand a formulated theory it may use empirical data. Basic research involves systematic, highly sophisticated scientific techniques. Fundamental research may not suggest the solutions of immediate problems, it rather draws long term conclusions.

9. Exploratory Research.

Ans :

Meaning

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

10. Variable.

Ans :

Meaning

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.

Choose the Correct Answers

1. Who was the author of the book named "Methods in Social Research"? [c]
   (a) Kerlinger (b) CR Kothari
   (c) Goode and Hatt (d) Wilkinson
2. Which of the following features are considered as critical in qualitative research? [c]
   (a) Collecting data with the help of standardized research tools.
   (b) Design sampling with probability sample techniques.
   (c) Collecting data with bottom-up empirical evidence.
   (d) Gathering data with top-down schematic evidence.
3. How is random sampling helpful? [d]
   (a) Reasonably accurate (b) An economical method of data collection
   (c) Free from personal biases (d) All of the above
4. A research intends to explore the result of possible factors for the organization of effective mid-day meal interventions. Which research method will be most appropriate for this study? [c]
   (a) Descriptive survey method (b) Historical method
   (c) Ex-post facto method (d) Experimental method
5. How to judge the depth of any research? [c]
   (a) By research title (b) By research duration
   (c) By research objectives (d) By total expenditure on research
6. Which of the following is not the method of Research? [c]
   (a) Survey (b) Historical
   (c) Observation (d) Philosophical
7. Research is [c]
   (a) Searching again and again
   (b) Finding solution to any problem
   (c) Working in a scientific way to search for truth of any problem
   (d) None of the above
8. In the process of conducting research ‘Formulation of Hypothesis” is followed by [c]
   (a) Statement of Objectives (b) Analysis of Data
   (c) Selection of Research Tools (d) Collection of Data
9. The main objective of _______ study’s to acquire knowledge [b]
   (a) Exploratory (b) Descriptive
   (c) Diagnostic (d) Descriptive and Diagnostic
10. _______ is concerned with discovering and testing certain variables with respect to their association or disassociation [c]
    (a) Exploratory (b) Descriptive
    (c) Diagnostic (d) Descriptive and diagnostic

Fill in the Blanks

1. ________ is systematized effort to gain new knowledge.
2. ________ research aims at discovery of new facts and verifications of old ones.
3. ________ research may be defined as research activities carried out relating to the different functionaries in the business and corporate world.
4. ________ research involves the develop- ment of new theories, abstract ideas, and generalized principles.
5. A ________ can be simply defined as any factor that varies.
6. ________ is the assignment of numerals to objects or events according to rules.
7. ________ classification is the property of a measure which is used to differentiate objects or its characteristics into various categories.
8. Nominal ________ is simply a system of assigning number symbols to events in order to label them.
9. ________ scales have an absolute or true zero of measurement.
10. The faculty member who is accountable for assigning the topic for the study is known as the ________

ANSWERS

1. Research
2. Social science
3. Business
4. Conceptual
5. Variable
6. Measurement
7. Distinctives
8. Scale
9. Ratio
10. Sponsor

One Mark Answers

1. Research.

Ans :

Research is an Organized and Systematic way of Finding answers to Questions.

2. Qualitative research.

Ans :

Qualitative research, on the other hand, is based on qualitative variables, which vary in quality of type.

3. Research design.

Ans :

A research design is a well-defined plan of action. It is a planned sequence of the entire research process. It is a blue print of research activity.

4. Independent variables.

Ans :

The variables which are supposed to be responsible for bringing about change(s) in a phenomenon are called independent variables.

5. Scaling.

Ans :

Scaling is an advancement of the measuring concept that refers to the continuum on which, the objects to be measured are marked.

UNIT II	**Research Design:** Research Problem, Purpose of Research Design, Types of Research Design: Experimental Research Design, Research Design for Cross Sectional, Longitudinal Studies, Characteristics of Good Research Design, Sampling and its Applications. Data Collection Methods & Tools: Types of Data, Sources and Instruments for Data, Guidelines for Questionnaire, Sampling and its Application. Measurement and Scaling, Reliability and Validity in Measurement of Variables, Sources of Error in Measurement.

2.1 Research Problem

Q1. What are Research Problems?

(OR)

What is a research problem? Define the main issues, which should receive the attention of the researcher in formulating the research problem. Give suitable examples to elucidate your points.

Ans : **(March-22)**

Meaning

Research is a scientific, systematic and purposeful search, for new knowledge or for re-interpretation of existing knowledge. It is a journey, which starts with a problem and ends with a solution. Identifying a research problem is the first and foremost step in a research process. The statement of research problem is the axis around which the whole research revolves, because it explains in brief the aims and objectives of the research.

A research problem is a specific statement in the general area of investigation. It is a precise identification of a problem situation in a certain context involving what, why, who, where and when of the problem area.

- **Who –** means the person or business organization that is facing a problem.
- **Why –** means that there is a purpose, goal aim or objective to solve this problem.
- **How –** means the options of actions one can take to solve the problem.
- **When –** means the time frame in which the problem is to be solved.
- **Where –** means the environment in which the problem exists.
- **What –** means the optimum action that is to be taken in solving the problem to attain the best results.

Selecting a Research Problem

1. **Sources of problems**

   First of all one should look at the sources from which one can select a research problem. Those may be readily available problem that has been identified by a person or an organization. If that is not the case, one can make use of the experience of experts of that field. The survey of related literature may also help in selection of a research problem.

2. **Potential to be a research problem**

   One must ensure that the problem one has undertaken has potential to be called as a research problem. One should avoid trivial or meaningless problem. A research problem must provide solution to an existing problem or contribute to the body of knowledge.

3. **Select non-controversial issues**

   Unless the problem is specifically related to issues such as religion, dogmas, beliefs, sexual preferences etc; one should try to avoid taking up controversial subjects.

4. **Researcher's interest and competency**

   In selection of research problem a researcher should choose a topic of his own field of study in which he has independent mastery in both the subject and method. The problem should sustain his interest, stimulate his imagination and should be within the range of his competencies.

5. Resources available

To conduct a research study various resources are needed.

A researcher must ensure the following points about the inputs of the problem:

(a) Time: The study must be completed in the allotted time frame.

(b) Funds: The amount of funds available from the sponsoring agencies must be known in advance and the study must be completed within given budget.

(c) Size of research: The size of the investigation must be manageable and should not be too large to handle. It should also not be too small to appear as a trivial problem.

(d) Co-operation of others: A researcher must ensure the necessary cooperation of colleagues and operational help of administrative authorities is available to the problem he is selecting.

(e) Literature and Material: Related literature is readily available to compare and support the research findings.

(f) Obtainable data: The selection of a problem should be such that the information or data needed for it is either readily available or obtainable.

Q2. Explain necessity of defining a Research Problems?

Ans :

Research is a disciplined approach to inquiry. It is a tool for testing as well as generating theories. Every research plan is unique in itself and has its unique research problems. An old saying is, "a problem well defined is a problem half solved".

By defining a research problem we mean a systematic way of asking and answering research questions. Meaningful, productive and useful research depends on the development of appropriate research questions, the identification of appropriate constructs and confidence we can have in our findings.

There are two main steps in defining a research problem:

1. Formulation of the problem
2. Establishment of research objectives

1. Formulation of the problem

Formulation of a problem is the most important step in a research process.

A clear statement of the problem is a key to good research. A firm may spend hundreds or thousands of rupees conducting research, but if it has not correctly identified the problem, this money is wasted. A clearly formulated research problem must answer all questions of the type Who, Why, How, When, Where, and What, regarding a research study.

2. Establishment of research objectives

With the problem or opportunity defined, the next step is to set objectives for research operations. Clear objectives can lead to clear results. Research objectives, related to and determined by the problem formulation, are set so that when achieved they provide the necessary information to solve the problem. A good way of setting research objectives is to ask, "What information is needed in order to solve the problem?" Your objective might be to explore the nature of a problem so you may further define it, or perhaps it is to determine how many people will buy your product packaged in a certain way and offered at a certain price. Your objective might even be to test possible cause and effect relationships.

For example, if you lower your price, how much will it increase your sales volume? And what impact will it have on your profit?

The problem description, the research question, sub questions and the research objectives are part of an overall definition of a research problem.

Q3. What are the sources of research problem ?

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

(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.

Q4. Briefly explain Research Problems in the fields of marketing, production, finance and personnel management.

Ans : **(May-22)**

1. Marketing

This is one area of business where research is the lifeline and is carried out on a vast array of topics and is conducted both in-house by the organization itself and out sourced to external agencies. Broader industry-or product-category-specific studies are also carried out by market research agencies and sold as reports for assisting in business decisions. Studies like these could be:

- Market potential analysis; market segmentation analysis and demand estimation.
- Market structure analysis which includes market size, players and market share of the key players.
- Sales and retail audits of product categories by players and regions as well as national sales; consumer and business trend analysis sometimes including short-and long-term forecasting.

2. Personnel and Human Resource Management

Human Resources (HR) and organizational behaviour is an area which involves basic or fundamental research as a lot of academic, macro-level research may be adapted and implemented by organizations into their policies and programmes.

Applied HR research by contrast is more predictive and solution-oriented. Though there are a number of academic and organizational areas in which research is conducted, yet some key contemporary areas which seem to attract more research are as follows:

- **Performance Management**

  This includes leadership analysis development and evaluation; organizational climate and work environment studies; talent and aptitude analysis and management; organizational change implementation, management and effectiveness analysis.

- **Employee Selection and Staffing**

  This includes pre and on-the-job employee assessment and analysis; staffing studies.

- **Organizational Planning and Development**

  This includes culture assessment either organization specific or the study of individual and merged culture analysis for mergers and acquisitions; manpower planning and development.

- **Incentive and Benefit Studies**

  These include job analysis and performance appraisal studies; recognition and reward studies, hierarchical compensation analysis; employee benefits and reward analysis, both within the organization and industry best practices.

3. Financial and Accounting Research

The area of financial and accounting research is so vast that it is difficult to provide a pen sketch of the research areas.

In this section, we are providing just a brief overview of some research topics:

- **Asset Pricing, Corporate Finance and Capital Markets**

  The focus here is on stock market response to corporate actions (IPOs or Initial Public Offerings, takeovers and mergers), financial reporting (earnings and firm specific announcements) and the impact of factors on returns, e.g., liquidity and volume.

- **Financial Derivatives and Interest Rate and Credit Risk Modeling**

  This includes analysing interest rate derivatives, development and validation of corporate credit rating models and associated derivatives; analysing corporate decision-making and investment risk appraisal.

- **Market Based Accounting Research**

  This includes analysis of corporate financial reporting behaviour; accounting-based valuations; evaluation and usage of accounting information by investors and evaluation of management compensation schemes.

- **Auditing and Accountability**

  This includes both private and public sector accounting studies, analysis of audit regulations; analysis of different audit methodologies; governance and accountability of audit committees.

Production and Operation Management

This area of management is one in which quantifiable implementation of the research results takes on huge cost and process implications. Research in this area is highly focused and problem specific. The decision areas in which research studies are carried out are as follows:

- Operation planning which includes product/service design and development; resource allocation and capacity planning.
- Demand forecasting and decision analysis.
- Process planning which includes production scheduling and material requirement management; work design planning and monitoring introduction scheduling and material requirement management; work design planning and monitoring.
- Project management and maintenance management studies.

2.2 Research Design

Q5. Give various definitions of research design explaining its meaning.

(OR)

What is a research design?

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

Meaning

A research design is a controlling plan for a research study in which the methods and procedures for collecting and analyzing the information to be collected is specified. It is a framework or plan for study that guides the collection and analysis of data.

The word 'design' means to work out the structure of form', as by making a sketch or plan. Thus, 'Research Design' is planning a strategy or drawing a blue print of conducting research. It is a guideline for collecting and utilizing data so that desired information can be obtained with sufficient precision and hypothesiscan be tested properly. A research is designed for the purpose of producing results that may be applied to real world situations. It not only enables a researcher to anticipate potential problems that can occur during the actual operation of the research, but also to limit boundaries of research study.

Definitions

Some of the popular definitions of research design are:

1. **According to By Miller** "Research design is the planned sequence of the entire process involved in conducting a research study."

2. **According to By Selltiz and others** "Research design is a catalogue of the various phases and facts relating to the formulation of a research effort. It is an arrangement of the essential conditions for collection and analysis of data in a form that aims to combine relevance to research purpose with economy in the procedure".

3. **According to By Anonymous** "A research designates the logical manner in which individuals or other units are compared and analyzed, it is the basis of making interpretations from the data".

   In short, research design is a plan of what data to gather, from whom, how and when to collect the data, and how to analyze the data obtained.

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

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.

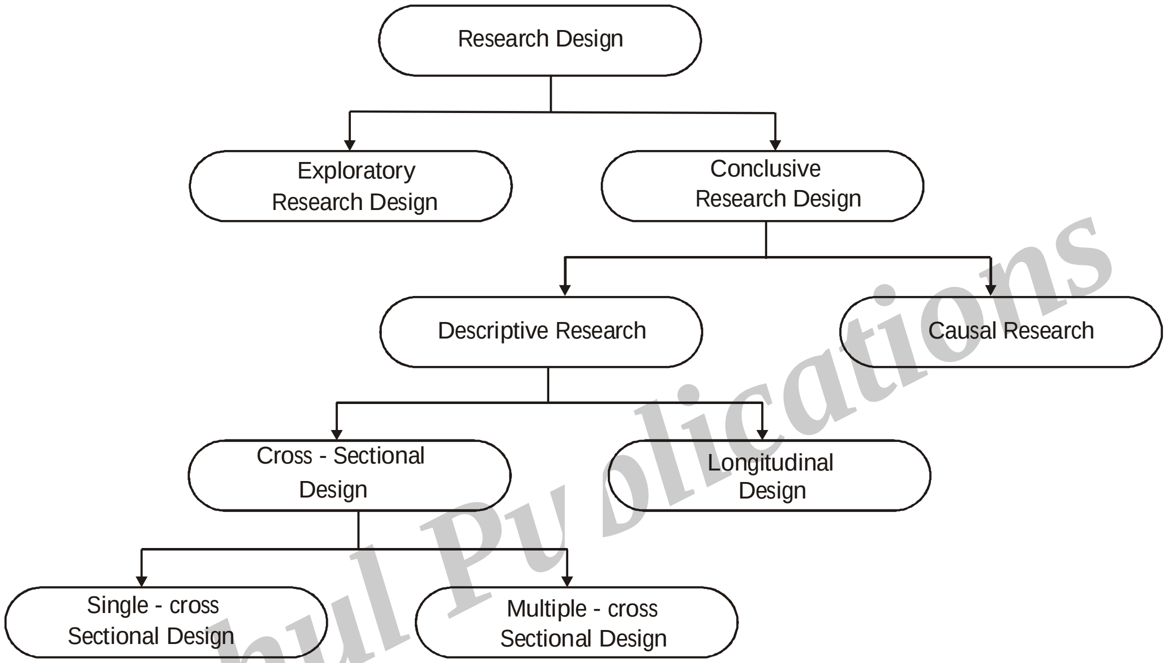
2.3 Types of Research Design

2.3.1 Experimental Research Design, Research Design for Cross Sectional, Longitudinal Studies

Q7. Discuss the different types of common research design.

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

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

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

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

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

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

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

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

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

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

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.

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

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 re-interviewed from time to time. Generally, panel data is related to the measurements of the same variables repeatedly.

Advantages

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

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:

i) The Method of Agreement
ii) The Method of Negative Agreement
iii) The Method of Concomitant Variation

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.

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 effect of varying more than one factor are to be determined. They are especially important in 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

	Experimental Variable	
Control variables	Treatment A	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

Q9. What are the advantages and disadvantages associated with cross-tabulation as a method of presenting and analysis of data ?

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

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.

Q10. Explain "Before-after experimental design with control group".

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

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.

Conditions	**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.

2.4 Characteristics of Good Research Design

Q11. What are the features of Good Research Design?

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

The following are the Characteristics of Good Research Design

1. Freedom from bias

A good research design should ensure that the method of data collection and analysis would not cause the data to vary in a systematic way. That is to say that the data should be free from systematic errors.

2. Freedom from confounding

In a good research design the variables involved in the study are separated from each other so that they do not influence each other.

3. Control of extraneous variables

In a well-designed research study the variables that are not under scrutiny do not influence the experimental variables in a systematic way for example things like temperature, time of day etc.

4. Statistical precision for testing hypothesis

A research design should ensure that the data are recorded at a level of precision that will yield statistically meaningful results.

5. With in resources

A design should draw limits of a research study so that it could be completed within available resources like time, money and staff.

6. Optimality

The best research design is one, which yields maximum precision in terms of bias and variance using minimum resources in terms sample size, time and money.

7. Objectivity

If operated by more than one researcher a good research design obtains same results. Thus, a good research design should be free from the subjectivity of its performer.

8. Flexibility

It is often observed that one has to deviate from the basic research design during the operation of the research study due to real world problems. A good research design is one, which not only has the potential to predict such practical problems, but also is flexible enough to incorporate changes in it whenever needed.

2.5 Sampling

Q12. What is Sampling?

Ans :

Meaning

A Sampling is a part of the total population. It can be an individual element or a group of elements selected from the population. Although it is a subset, it is representative of the population and suitable for research in terms of cost, convenience and time. The sample group can be selected based a probability or a non-probability approach. A sample usually consists of various units of the population. The size of the sample is represented by TT.

Sampling is the act, process, or technique of selecting a representative part of a population for the purpose of determining the characteristics of the whole population. In other words, the process of selecting a sample from a population using special sampling techniques called sampling. It should be ensured in the sampling process itself that the sample selected is representative of the population.

Though the sampling is not new but sampling theory has been developed recently. People know or not but they have been using the sampling technique in their day to day life.

For example:

A house wife tests a small quantity of rice or wheat to see whether it is of a good quality and gives the generalized result about the whole rice kept in the bag or vessel. The result arrived at most of the times is 100% correct.

Q13. Define the following terms :

(i) Population

(ii) Sample

Ans :

(i) Population

In a statistical investigation the interest usually lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristics relating to individuals belonging to a group. This group of individuals under study is called as population or universe.

So. population is defined as, "an aggregate of objects, animate or inanimate under study." It may be finite or infinite.

(ii) Sample

A finite subset of statistical individuals in a population is called sample.

It is quite often used in our day to day life.

For example: Assessing the quality of foodgrain by taking a handful of it from the bag.

Q14. Explain the advantages of sampling.

Ans :

Advantages

1. It saves time

Sampling method of data collection saves time because fewer items are collected and processed. When the results are urgently required, this method is very useful.

2. It reduces cost

Since only a few and selected items are studied in sampling, so there is a reduction in cost of money and reduction in terms of man-hours.

3. More reliable results

Through sampling more reliable results can be obtained because -

(a) There are fewer chances of sampling statistical errors. Though, any sampling error, it is possible to estimate and control the results.

(b) Highly experienced and trained persons can be: employed for scientific processing and analyzing of relatively limited data and they can use their high technical knowledge and get more accurate and reliable results.

4. It provides more detailed information

As it saves time, money and labour more detailed information can be collected in a sample survey.

5. Sometimes only method to depend upon

It is observed that sometimes one has to depend only upon sampling method. It happens when the population under study is infinite. In such situation it is only the method to be used.

For example: If someones blood has to be examined, it will become fatal to take all the blood out from the body and study depending upon the total enumeration method.

6. Administrative convenience

The organization and administration of sample survey are easy for the same time, money and labour reasons.

7. More scientific

Since the methods used to collect data are based on scientific theory and results obtained can be tested, sampling is a more scientific method to collect data.

Q15. Discuss the techniques of sampling.

(OR)

Briefly explain random and non-random sampling.

Ans : **(March-22)**

Methods (or) Techniques of Sampling

There are two basic approaches to sampling: probabilistic and non-probabilistic sampling. Let us look at the various types of sampling under each category:

1) Probability Sampling

i) Simple Random Sampling

ii) Systematic Sampling

iii) Stratified Sampling

iv) Multistage Cluster Sampling

2. Non-probability Sampling

i) Convenience Sampling

ii) Quota Sampling

iii) Judgment Sampling

iv) Snowball Sampling

1. Probability Sampling

A sampling in which every member of the population has a calculable and non-zero probability of being included in the sample is known as probability sampling. Methods of random selection consistent with both the probabilities of inclusion are used in forming estimates from the sample. The probability of selection need not be equal for members of the population. If the purpose of a research is to arrive at conclusions or make predictions affecting the population as a whole, then the choice of a probabilistic sampling approach is desirable.

(i) Simple Random Sampling

A sampling process where each element in the target population has an equal chance or probability of inclusion in the sample is known as simple random sampling. For example, if a sample of 15,000 names is to be drawn from the telephone directory, then there is equal chance for each number in the directory to be selected. These numbers (serial no. of names) could be randomly generated by the computer or picked out of a box. These numbers could be later matched with the corresponding names thus fulfilling the list. In small populations random sampling is done without replacement to avoid the instance of a unit being sampled more than once.

The benefits of simple random sampling can be reaped when the target population size is small, homogeneous, sampling frame is clearly defined, and not much information is available regarding the population. It is advantageous in that it is free of classification error, and requires minimum advance knowledge of the population. Two striking features are the elimination of human bias and non- dependency on the availability of the element. It is seldom put into practice because of the application problem associated with it. This sampling method is generally not preferred as it becomes imperative to list every item in the population prior to the sampling and requires constructing a very large sampling frame, resulting in extensive sampling calculations and excessive costs.

(ii) Systematic Sampling

Systematic sampling involves the selection of every kth element from a sampling frame. Here 'k' represents the skip interval and is calculated using the following formulae.

Skip interval (k) = Population size/Sample size

Often used as a substitute to simple random sample, it involves the selection of units from a list using a skip interval (k) so that every k' th element on the list, following a random start between 1 and k. is included in the sample. For example, if k were to equal 6, and the random start were 2, then the sample would consist of 2^{nd}. 8^{th}, 14^{th}. 20^{th} elements of the sampling frame.

It is to be noted here that if the skip interval is not a whole number then it is rounded off to the nearest whole number. This sampling method can be used in industrial operations where the equipment and machinery in the production line are checked for proper functioning as per the specifications. The manufacturer can select every k' th item to ensure consistent quality or for detection of defects. Therefore, he requires the first item to be selected randomly at the starting point and subsequently he can choose every k'th item for evaluation against specifications. It also finds its applicability while questioning people in a sample survey where the interviewer may catch hold of every 10th person entering a particular shop. However, in every case, the researcher has to determine the skip interval and proceed thereafter. In both the cases, it is necessary to select the first item in the population in a random manner and thereafter follow the skip interval. This method is more economical and less time consuming than simple random sampling.

(iii) Stratified Random Sampling

Stratification is the process of grouping the members of the population in homogeneous group before sampling. It should be ensured that each element in the population is assigned a particular stratum only. The random sampling is applied within each stratum independently. This often improves the representativeness of the sample by reducing the sampling error.

The number of units drawn for sampling from each stratum depends on the homogeneity of the elements. A smaller sample can be drawn from the known to have the elements with the same value whereas sample can be drawn in much higher proportion from another stratum where the values are known to differ. This is because in the former case the information from the smaller number of respondents can be enumerated to the whole sample stratum However in the latter case with much variability among the elements the higher elements value will keep the sampling to minimum errors to minimum value. The smaller errors may be due to groups are appreciably represented when strata are combined.

(iv) Multistage Cluster Sampling

Clustering involves grouping the population into various clusters and selecting few clusters for study. Cluster sampling is suitable for conducting research studies that cover large geographic area.

Once the cluster is formed the researcher can either go for one stage, two stages, or multistage cluster sampling. In single stage, all the elements from each selected are studied, whereas in two stages, the researchers use random to select few elements from clusters. Multistage sampling involves selecting a sample in two or more successive stages. Here the cluster selected in the first stage can be divided into cluster units.

For example consider the case where a company decides to interview 400 households about the likeability of its new detergent in a metropolitan city. To minimize the resources and time researchers divide the city into separate blocks say 40, each block consist of heterogeneous units. The researcher may opt for the two stage cluster sampling if he finds that individual clusters have little heterogeneity to other clusters. Similarly a multistage cluster sampling involves three or more sampling steps, it differs from stratified sampling that is done in cluster in contrast to elements within strata as is the case in the stratified sampling. Elements are randomly selected from each stratum in case of stratified sampling whereas only selected clusters are studied in cluster sampling.

2. Non-probability Sampling

It involves the selection of units based on factors other than random chance. It is also known as deliberate sampling and purposive sampling. For example, a scheme whereby units are selected purposefully would yield a non-random sample. In a general sense, it is an umbrella term which includes any sample that does not conform to the requirements of a probability sampling. Convenience sampling, quota sampling, judgment sampling and snowball sampling are few examples of non-probability sampling.

Methods of Sampling

Probability Sampling

- Sample Random Sampling
- Systematic Sampling
- Stratified Random Sampling
- Multistage Cluster Sampling

Non-probability Sampling

- Convenience Sampling
- Quota Sampling
- Judgement Sampling
- Snowball Sampling

Fig.: Methods of Sampling

(i) Convenience Sampling

The selection of units from the population based on their easy availability and accessibility to the researcher is known as convenience sampling. For example, imagine a Co., that surveys a sample of its employees to know the acceptance for a new flavor of potato chips that it plans to introduce in the market. This type of sampling is a typical example of convenience sampling as the criterion for selecting a sample is convenience and availability. Although this type of research is easy and cost effective, the findings of the sample survey cannot be generalized to the entire population, as the sample is not representative. As there is no set criterion for selecting the sample, there is a scope for research being influenced by the bias of the researcher. As in the above ex, the researcher may conduct a sample survey involving its own employees to find whether the market, would accept the product.

(ii) Quota Sampling

In quota sampling, the entire population is segmented into mutually exclusive groups. The number of respondents (quota) that are to be drawn from each of several categories is specified in advance and the final selection of respondents is left to the interviewer who proceeds until the quota for each category is filled. Quota sampling finds extensive use in commercial research where the main objective is to ensure that the sample represents in relative proportion, the people in the various categories in the population, such as gender, age group, social class, ethnicity and region of residence. For example, if a researcher wants to segment the entire population based on gender, then he would have two categories of respondents, that is, males and females. If he plans to collect a sample of 30, he may allot a quota of 15 for male and 15 for female respondents. Therefore, the researcher will stop administering the questionnaire to females after he interviews the 15th female respondent, that is, when the quota of 15 females is filled.

Quota sampling is subject to interviewer bias that may result in:

- The quota reflecting the population in terms of superficial characteristics.
- The researcher selecting the respondents based on availability rather than on their suitability to the study.

(iii) Judgment Sampling

The selection of a unit, from the population based on the judgment of an experienced researcher, is known as judgment or purposive sampling. Here, the sample units are selected based on population's parameters. It is often noticed that companies frequently select certain preferred cities during test marketing their products. This is because they consider the population of that particular city to be representative of the total population of the country. The same is the case with the selection of specific shopping malls that according to the researcher's judgment attract a reasonable number of customers from different sections of the society. Polling results predicted on television is also a result of judgment sampling. Researchers select those districts that have voting patterns close to the overall state or country in the previous year. The judgment of the researcher is based on the assumption that the past voting trends of selected sample districts are still representative of the political behavior of the state's population. For example, certain companies test market their new product launches in cities like Mumbai and Bangalore, because the profile of these cities is representative of the total Indian population.

(iv) Snowball Sampling

Sampling procedures that involve the selection of additional respondents are known as snowball sampling. This sampling technique is used against low incidence or rare populations. Sampling is a big problem in this case, as the defined population from which the sample can be drawn is not available. Therefore, the process sampling depends on the chain system of referrals. Suppose. SG sports Ltd.. a manufacturer of sports equipment plans to survey 100 senior players through its new website for getting their feedback on the quality of its products.

However, keeping track of such senior senior squash players can be very difficult, as their presence may be very rare or low. Therefore, it collects the details of the first 200 visitors to its website, to list if any of them is a squash player or knows a squash player. If the visitor is a squash player, then he is requested to refer the names of at least 3 other players known to him. The referred names of the squash players are then called upon for further referrals and this goes on until the sample size of 100 adult players is reached. Although small sample sizes and low costs are the clear advantages of snowball sampling, bias is one of its disadvantages. The referral names obtained from those sampled in the initial stages may be similar to those initially sampled. Therefore, the sample may not represent a cross-section of the total population. It may also happen that visitors to the site or interviewers may refuse to disclose the names of those whom they know.

Q16. Explain the Essentials of Sampling.

Ans :

Principles

In order to reach to the clear conclusion, sampling should possess following essentials:

1. It must be representative

The sample selected should possess the similar characteristics of the original universe from which it has been drawn.

2. Homogeneity

Selected samples from the universe should have similar nature and should not have any difference when compared with the universe.

3. Adequate samples

In order to have a more reliable and representative results, a good number of items are to be included in the sample.

4. Optimization

All efforts should be made to get maximum results both in terms of cost as well as efficiency.

If size of the sample is larger, there is better efficiency and at the same time cost is more. A proper size of sample is maintained in order to have optimized results in terms of cost and efficiency.

Q17. What is Bias in Survey ? What are the reasons for biases? Explain the same briefly.

Ans : **(May-22)**

A personal and sometimes unreasonable judgment for or against someone or something. If someone is biased, they prefer a group of people, an idea, or a thing over another, often due to personal and unreasoned judgment, and behave unfairly as a result.

1. Leading Questions

A leading question “leads” the respondent toward a “correct” answer by wording questions in a way that sways readers to one side. For example, a survey may ask, “How dumb are the President’s policies?”, which puts it into the respondent’s head that, yes, the policies are dumb. A non-leading question would be, “What do you think of the President’s policies?”

Leading questions can also happen on scaled questions. For example, a satisfaction survey could ask participants if they were extremely satisfied, satisfied, or dissatisfied. By providing more options on the satisfied side, the question is biased and leading participants in that direction.

2. Loaded Questions

These questions more or less force respondents to answer a question in a particular way. For example: “Where do you like to go on weekends?” Well, what if a respondent likes to stay home on the weekend? A better, unloaded question would be, “What do you like to do on weekends?” You can read more in our blog on leading and loaded questions.

3. Double-Barreled Questions

Researchers sometimes make the mistake of putting two questions in one. For example, “Do you like your boss and his policies?” This poses a problem if a respondent likes their boss as a person, but does not like his policies. Instead, the question should be broken into two separate questions.

4. Absolute Questions

Otherwise known as the “yes or no” question. For example, if you ask “do you always brush your teeth before bed,” you’re backing respondents into a corner. While most may brush before bed, certainly there are times they forget or are too tired, so they may answer negatively even though 9 out of 10 times the answer is yes. A likert scale with options can solve this issue.

5. Unclear Questions

Clarity is key when developing surveys, so avoid overly technical terminology, jargon, abbreviations, or acronyms that people may not understand. There are also times when you don’t want to be too specific; for example, if you’re conducting a survey about smartphone use, don’t use “iPhone” in your questions just because they may be the most popular.

2.5.1 Applications

Q18. 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.

2.6 Data Collection Methods & Tools

Q19. Define :

(i) Data

(ii) Information

Ans :

(i) Data

Data can exist in a variety of forms as numbers or text on pieces of paper, as bits and bytes stored in electronic memory, or as facts stored in a person's mind. Strictly speaking, data is the plural of datum, a single piece of information. In practice, however, people use data as both the singular and plural form of the word.

Data that is accurate and timely, specific and organized for a purpose, presented within a context that gives it meaning and relevance, and can lead to an increase in understanding and decrease in uncertainty.

(ii) Information

Information in raw or unorganized forms (such as alphabets, numbers, or symbols) that refers to, or represent, conditions, ideas or objects. Information is valuable because it can affect behaviour, a decision or an outcome. For example, if a manager is told his/her company's net profit decreased in the past month, he/she may use this information as a reason to cut their financial spending.

Data vs Information

Data is raw, unorganized facts that need to be processed. Data can be something simple and seemingly random and useless until it is organized. Whereas when data is processed, organized, structured or presented in a given context so as to make it useful, is called Information.

Each student's test score is one piece of data. Whereas the classes, average score or the school's average score is the information that can be concluded from the given data.

The terms data, information and knowledge are frequently used for overlapping concepts. The main difference is in the level of abstraction being considered. Data is the lowest level of abstraction, information is the next level, and finally, knowledge is the highest level among all three. Data on its own carries no meaning. For data to become information, it must be interpreted and take on a meaning. For example, the height of Mt. Everest is generally considered as "data", a book on Mt. Everest geological characteristics may be considered as "information", and a report containing practical information on the best way to reach Mt. Everest's peak may be considered as "knowledge".

2.6.1 Types of Data

Q20. Explain different types of data.

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

There are two types of data :

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.

2.6.2 Sources and Instruments for Data

Q21. Explain the various methods of collecting primary data.

(OR)

List and describe some sources of primary data collection

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

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.

Q22. What are the advantages and disadvantages of primary data.

Ans :

Advantages

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 organization 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

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.

Q23. 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.

Q24. Explain the various methods of collecting secondary data.

(OR)

What are the methods for collecting secondary data?

Ans : **(March-22, Nov.-21)**

Meaning

Secondary data are those data which have been already collected and analyzed 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 organization. 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.

Definition

(i) 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".

Purpose

- 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 organizations, 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 Organizations

International organizations such as the International Monetary Fund, the World Bank, the United Nations Organization, 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 organizations, 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 organizations in foreign countries, which are selling their research findings to organizations requiring knowing the outcome of their researches.

Q25. What are the advantages and dis-advantages of Secondary Data?

(OR)

What are the Limitations of Secondary data ?

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

Advantages

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 specialized in data collection. Also many organizations 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

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.

Q26. Explain the Various ways Editing of Secondary Data.

Ans :

1. Test of Reliability

While editing the secondary data, the editor must see that the data obtained are accurate and reliable.

For testing the reliability of the data, the editor should make the following queries:

i) Who collected the data ?

ii) Where from the data were collected ?

iii) Is the compiler dependable in regard to honesty, integrity, experience and training ?

iv) Is the source of the data dependable in regard to accuracy, adequacy and consistency ?

v) What methods were employed in the primary collection of the data ?

vi) Are the methods of collection proper and dependable ?

vii) At what time, the data were collected and was it a normal time ?

viii) Was there any possibility of bias, and prejudices creeping into the minds of the compilers ?

ix) What degree of accuracy was fixed by the investigator and was it achieved ?

x) Was the size of the sample adequate ?

xi) Was the sample at random, or adequate ?

xii) With what purpose the data were collected ?

xiii) What period is covered by the data and how far it is relevant for the present study ?

xiv) What units of collection and measurement were employed ? Were they clearly defined ? Are they suitable for the present purpose ?

xv) Were the editing, tabulation, and analysis of the data carefully and consciously done ?

2. Test of Adequacy

While editing the secondary data it must be seen that they are adequate, or sufficient for the purpose of the enquiry. As pointed out earlier, too much of data may prove to be confusing and irrelevant. Similarly, too less of data, also, will not serve the purpbse and give the true picture of the problem under study. Therefore, the data must be adequate for the purpose. Whether the data collected from the secondary sources are adequate, or not can be tested in the light of the following queries:

i) What was the geographical area from which the data were collected ?

ii) Is the area of collection wider or narrower than the area covered under the present study ?

For example, if the object is to measure the change in the general price level of India through the construction of a whole-sale price index number but the data collected relate only to the cost of living of the people in a particular locality, it would not serve the purpose on the ground of inadequacy.

iii) What is the period covered by the data ?

iv) Is the period covered by the data commensurate with the period of the problem under study ?

v) What was the degree of accuracy achieved with the collected data ?

vi) Is the degree of accuracy achieved with the data commensurate with the degree of accuracy desire in the present enquiry ?

For example, if in the collected data, 95% degree of accuracy was achieved, and in the present study, 99% degree of accuracy is desired, the data thus collected will not suit the purpose of the present enquiry.

3. Test of Suitability

While editing the secondary data, it must be seen that the data collected are suitable for the present study. If the data collected are not suitable, it will vitiate the whole purpose of the enquiry and lead to erroneous conclusions. The suitability of the data can be tested in the light of the following queries:

i) What was the nature of the problem for which the data were collected? If the nature of the problem under study does not resemble with that of the problem for which the data were originally collected, the same data will not be suitable for the investigation.

ii) What was the object of the enquiry ?

If the object of the past enquiry is completely different from that of the present enquiry, the collected data will not suit the present purpose. Thus, if the object of the present investigation is to study the trend in the wholesale price, but the data collected were for studying retail prices, such data would be unsuitable for the purpose.

iii) What was the scope of the enquiry?

If the geographical area covered by the data collected was more wider, If the geographical area covered by the data were collected to study the functioning of the non-banking financial institutes in New York, the said data will not be suitable for studying the same problem with reference to the state of Washington.

iv) Do the definitions given to the various terms and units used in the earlier investigation remain the same as those under the present enquiry ? If the definition of the various terms and the units used in the two enquiries are completely different, the data collected originally will not suit the present enquiry. For instance, if the term ‘wage’ under the present enquiry relates to the unskilled labour, the said collected data will not suit the present enquiry.

v) What was the time covered by the data in the earlier enquiry ?

If the time covered by the data was radically different from the time required to be covered by the data under the present enquiry, the data concerned will not be suitable for the problem under study.

Q27. What are the difference between Primary and Secondary Data ?

Ans : **(Imp.)**

S.No	Nature	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 analyzed 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 analyzed 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, 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.

2.7 Questionnaire

Q28. Define Questionnaire. Explain different types of Questionnaire.

Ans : **(Imp.)**

Meaning

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 realize 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

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?"

Q29. What are the Components of Questionnaire.

Ans :

The following are the Components of Questionnaire.

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.

Q30. 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 Questionnaire

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 analyze 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.

2.7.1 Guidelines for Questionnaire

Q31. What are the Guidelines / Precautions for Questionnaire ?

Ans : **(March-22, 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'.

Q32. 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 standardize. 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 analyze 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.

Q33. 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) Qualifacts
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.

Q34. Distinguish between questionnaire and schedule for collection of data.

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

S.No.	Nature	Questionnaire	Schedule
1.	**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 answers.
2.	**Filled by**	Respondents	Enumerators
3.	**Response Rate**	Low	High
4.	**Coverage**	Large	Comparatively small
5.	**Cost**	Economical	Expensive
6.	**Respondent's Identity**	Not known	Known
7.	**Success relies on**	Quality of the questionnaire	Honesty and competence of the enumerator
8.	**Usage**	Only when the people are literate and cooperative.	Used on both literate and illiterate people

2.8 Measurement and Scaling

Q35. Define the following terms :

(i) Measurement

(ii) Scaling

Ans :

For answer refer Unit-I, Q.No. 18, 21

2.9 Reliability and Validity in Measurement of Variables

Q36. 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 standardization the conditions under which the measurement takes place i.e., we must ensure that external sources of variation such as boredom, fatigue, etc., are minimized 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.

Q37. Explain the test of validity of measurement.

(OR)

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

(OR)

Explain any four formats of Validity

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

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.

Q38. Distinguish between the terms 'reliability' and 'validity'.

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

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.

2.9.1 Sources of Error in Measurement

Q39. What are the various sources of errors in measurement.

Ans : **(Imp.)**

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 panting, inadequate space for replies, response choice omissions, etc. are a few things that make the measuring instrument defective and 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.

Short Question and Answers

1. What are Research Problems?

Ans :

Meaning

Research is a scientific, systematic and purposeful search, for new knowledge or for re-interpretation of existing knowledge. It is a journey, which starts with a problem and ends with a solution. Identifying a research problem is the first and foremost step in a research process. The statement of research problem is the axis around which the whole research revolves, because it explains in brief the aims and objectives of the research.

A research problem is a specific statement in the general area of investigation. It is a precise identification of a problem situation in a certain context involving what, why, who, where and when of the problem area.

- **Who –** means the person or business organization that is facing a problem.
- **Why –** means that there is a purpose, goal aim or objective to solve this problem.
- **How –** means the options of actions one can take to solve the problem.
- **When –** means the time frame in which the problem is to be solved.
- **Where –** means the environment in which the problem exists.
- **What –** means the optimum action that is to be taken in solving the problem to attain the best results.

2. What is a research design?

Ans :

Meaning

A research design is a controlling plan for a research study in which the methods and procedures for collecting and analyzing the information to be collected is specified. It is a framework or plan for study that guides the collection and analysis of data.

The word 'design' means to work out the structure of form', as by making a sketch or plan. Thus, 'Research Design' is planning a strategy or drawing a blue print of conducting research. It is a guideline for collecting and utilizing data so that desired information can be obtained with sufficient precision and hypothesiscan be tested properly. A research is designed for the purpose of producing results that may be applied to real world situations. It not only enables a researcher to anticipate potential problems that can occur during the actual operation of the research, but also to limit boundaries of research study.

3. Purpose of Research Design.

Ans :

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.

4. Exploratory Research Design.

Ans :

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.

5. What is Sampling?

Ans :

Meaning

A Sampling is a part of the total population. It can be an individual element or a group of elements selected from the population. Although it is a subset, it is representative of the population and suitable for research in terms of cost, convenience and time. The sample group can be selected based a probability or a non-probability approach. A sample usually consists of various units of the population. The size of the sample is represented by TT.

Sampling is the act, process, or technique of selecting a representative part of a population for the purpose of determining the characteristics of the whole population. In other words, the process of selecting a sample from a population using special sampling techniques called sampling. It should be ensured in the sampling process itself that the sample selected is representative of the population.

6. Population.

Ans :

In a statistical investigation the interest usually lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristics relating to individuals belonging to a group. This group of individuals under study is called as population or universe.

So. population is defined as, "an aggregate of objects, animate or inanimate under study." It may be finite or infinite.

7. Quota Sampling.

Ans :

In quota sampling, the entire population is segmented into mutually exclusive groups. The number of respondents (quota) that are to be drawn from each of several categories is specified in advance and the final selection of respondents is left to the interviewer who proceeds until the quota for each category is filled.

8. Data.

Ans :

Data can exist in a variety of forms as numbers or text on pieces of paper, as bits and bytes stored in electronic memory, or as facts stored in a person's mind. Strictly speaking, data is the plural of datum, a single piece of information. In practice, however, people use data as both the singular and plural form of the word.

Data that is accurate and timely, specific and organized for a purpose, presented within a context that gives it meaning and relevance, and can lead to an increase in understanding and decrease in uncertainty.

9. Information.

Ans :

Information in raw or unorganized forms (such as alphabets, numbers, or symbols) that refers to, or represent, conditions, ideas or objects. Information is valuable because it can affect behaviour, a decision or an outcome. For example, if a manager is told his/her company's net profit decreased in the past month, he/she may use this information as a reason to cut their financial spending.

10. Define Questionnaire.

Ans :

Meaning

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.

Choose the Correct Answer

1. A formal document that presents the research objectives, design of achieving these objectives, and the expected outcomes/deliverables of the study is called ______. [b]
   (a) Research design (b) Research proposal
   (c) Research hypothesis (d) Research report
2. Survey is a ______ Study. [b]
   (a) Descriptive (b) Fact finding
   (c) Analytical (d) Systematic
3. In a survey there is an enumerator and a ______. [b]
   (a) Guide (b) Respondent
   (c) Supervisor (d) Messenger
4. The first purpose of a survey is to ______. [d]
   (a) Description (b) Evaluation
   (c) Pration (d) Provide Information
5. Questions in which only two alternatives are possible is called ______. [b]
   (a) Multiple choice questions (b) Dichotomous questions
   (c) Open ended questions (d) Structured questions
6. What are the core elements of a Research Process? [d]
   (a) Introduction; Data Collection; Data Analysis; Conclusions and Recommendations
   (b) Executive Summary; Literature Review; Data Gathered; Conclusions; Bibliography
   (c) Research Plan; Research Data; Analysis; References
   (d) Introduction; Literature Review; Research Methodology; Results; Discussions and Conclusions
7. Identifying causes of a problem and possible solution to a problem is ______. [b]
   (a) Field Study (b) Diagnosis tic study
   (c) Action study (d) Pilot study
8. Second step in problem formulation is ______. [b]
   (a) Statement of the problem (b) Understanding the nature of the problem
   (c) Survey (d) Discussions
9. Questionnaire is filled by ______. [a]
   (a) Respondent (b) Everybody
   (c) Enumerator (d) None of the above
10. The main problem in questionnaire is ______. [c]
    (a) Accessible to Diverse Respondent
    (b) Greater Anonymity
    (c) Shows an inability of respondent to provide information
    (d) None of these

Fill in the Blanks

1. A ________ is a specific statement in the general area of investigation.
2. Formulation of a ________ is the most important step in a research process.
3. ________ research is characterized by a high degree of flexibility.
4. A ________ study is concerned with a sample of elements from a given population.
5. ________ designs are used in experiments where the effect of varying more than one factor are to be determined.
6. A ________ is a part of the total population. It can be an individual element or a group of elements selected from the population.
7. ________ sampling is suitable for conducting research studies that cover large geographic area.
8. ________ data is one, which is collected by the investigator himself for the purpose of a specific inquiry or study.
9. ________ are a form of questioning that is more rigid than interviews and that involve larger groups of people.
10. The ________ is an important tool for gathering primary data.

Answers

1. Research problem
2. Problem
3. Exploratory
4. Cross-Sectional
5. Factorial
6. Sampling
7. Cluster
8. Primary
9. Surveys
10. Questionnaire

One Mark Answers

1. Field Studies.

Ans :

Field studies are ex-post-facto scientific enquires that aim at finding the relations and interrelations among variables in a real setting.

2. Simple Random Sampling.

Ans :

A sampling process where each element in the target population has an equal chance or probability of inclusion in the sample

3. Snowball Sampling

Sampling procedures that involve the selection of additional respondents are known as snowball sampling.

4. Questionnaire.

Ans :

A questionnaire is a prepared set of questions (or measures) used by respondents or interviewers to record answers (data).

5. Content Validity.

Ans :

Content validity is the extent to which a measuring instrument provides adequate coverage of the topic under study.

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. Introduction to Statistics, Measurement of Central Tendency and Dispersion. (b) **Small Sample Tests**: t-Distribution, Properties and Applications, Testing for One and Two Means, Paired t-Test, Hypothesis Formulation and Testing.

3.1 Data Classification and Tabulation

Q1. Define Classification. Explain the requisites of Ideal Classification.

Ans :

Classification is the grouping of related facts into classes.

i) **According to Connor (1997)** "The process of arranging things in groups or classes according to their resemblances and affinities and gives expression to the unity of attributes that may subsist amongst a diversity of individuals". "Classification is the process of arranging data into sequences and groups according to their common characteristics or separating them into different but related parts".

ii) **According to Secristi** "The process of grouping large number of individual facts and observations on the basis of similarity among the items is called classification".

Requisites of Ideal Classification

(i) **It should be unambiguous:** There should be no uncertainty or ambiguity. Classes should be defined rigidly, so as to avoid any ambiguity. It should be flexible: The classification should be enough to accommodate change, amendment and inclusion in various classes in accordance with new situations.

(ii) **It should be homogeneous:** Units of each class should be homogeneous. All the units included in a class or group should be present according to the property on basis of which the classification was done. It should be suitable for the purpose: The composition of the class should be according to the purpose.

For example: To find out the economic condition of the persons, create classes on the basis of income.

(iii) **It should be stable**

Stability is necessary to make data comparable and to make out meaningful comparison of the results. This means that the classification of data set into different classes must be performed in a way, that whenever an investigation is carried out, there is no change in classes and so the results of the investigation can be compared easily.

iv) It should be exhaustive

Each and every item of data must belong to a particular class. An ideal classification is one that is free from any residual classes such as others or miscellaneous, as they do not state the characteristics clearly and completely.

v) It should be mutually exclusive

The classes should be mutually exclusive.

Q2. Explain different types of Classifications.

Ans :

The data can be classified on the basis of following four criteria:

a) Geographical Classification

b) Chronological Classification

c) Qualitative Classification

d) Quantitative Classification

a) Geographical Classification: When data are classified with reference to geographical locations such as countries, states, cities, districts etc, it is known as geographical classification.

For Example: The production of rice in different states of India, production of wheat in different countries etc.

State-wise comparison of Rice production in India

Name of State	Production of Rice (Metric tonnes/hectares)
Andra Pradesh	7.49
Bihar	6.5
Chhattisgarh	6.09
Punjab	11.82
Tamil Nadu	7.98
Utter Pradesh	12.5
West Bengal	15.75

Geographical Classification are usually listed in alphabetical order for easy reference. Items may also be listed by size to emphasis the important areas as in ranking the States by population.

b) Chronological Classification: Classification where data are grouped according to time is known as chronological classification

For Example the population of India from 1931 to 2001.

Population of India from 1931 to 2001

Year	Population (in millions)
1931	276
1941	313
1951	357
1961	438
1971	536
1981	634
1991	846
2001	1002

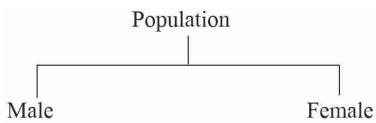
Time series are usually listed in chronological order, normally starting with the earliest period. When the major emphasis falls on the most recent events, a reverse time order may be used.

c) **Qualitative Classification:** In Qualitative classification, data are classified on the basis of some attributes or quality such as gender, religion, literacy, marital status etc. In this type of classification, the attribute under study cannot be measured. It can only be found out whether it is present or absent in the units of study.

For example, if the population to be classified in respect to one attribute, say gender, then we can classify them into two classes namely males and females.

Thus when only one attribute is studied two classes do formed, one possess the attribute and the other not possessing the attribute. This type of classification is known as simple or dichotomous classification.

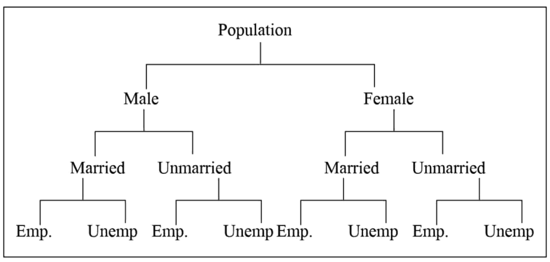
For example, if the population under study is divided into categories as follows:



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 gender, each of these classes may subdivided into married and unmarried on the basis of marital status. Further classification can be made on the basis of attribute say, employment.

Example of manifold classification is as follows:



(d) Quantitative Classification: It refers to the classification of data according to some characteristics that can be measured such as height, weight, profits, income, sales etc.

For example, the students of the school may be classified according to weight as follows:

Weight (in kg)	No. of Students
40-50	40
50-60	160
60-70	110
70-80	200
80-90	90
Total	**600**

Such a distribution is known as empirical frequency distribution or simple frequency distribution.

In this type of classification, there are two elements,

a) The variable (weight in above example)

b) The frequency (the number of students in each class) i.e. there are 40 students having weight ranging from 40 to 50 kg, 160 students having weight 50 to 60 kg and so on.

Thus, we can find out the ways in which the frequencies are distributed.

Q3. Explain various methods of data classification.

Ans :

There are two methods of classifying the data according to class intervals.

(a) Exclusive method: When the class intervals are so fixed that the upper limit of one class is the lower limit of the next class it is known as the exclusive method of classification.

Weight (in kg)	No. of Students
40-50	40
50-60	160
60-70	110
70-80	200
80-90	90

In the above example, there are 40 students whose weight is between 40 to 49.99 kg. A student whose weight is 50 kg is included in the class 50 – 60.

(b) Inclusive method: In this method, the upper limit of one class is included in that class itself.

Weight (in kg)	No. of Students
40-49	40
50-59	160
60-69	110
70-79	200
80-89	90

In the class 40 – 49, we include students whose weight is between 40 kg and 49 kg. If the weight of the student is exactly 50 kg he is included in next class.

Q4. Define tabulation. State the objectives of tabulation.

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

Meaning

The simplest and most revealing devices for summarizing data and presenting them in a meaningful manner is the statistical table. After classifying the statistical data, next step is to present them in the form of tables. A table is a systematic organization of statistical data in rows and columns. The purpose of a table is to simplify the presentation and to facilitate comparisons. The main objective of tabulation is to answer various queries concerning the investigation. Tables are very helpful for doing analysis and drawing inferences from them.

Classification and tabulation go together, classification being the first step in tabulation. Before the data are put in tabular form, they have to be classified.

Objectives

- **To simplify complex data:** It reduces raw data in a simplified and meaningful form. The reader gets a very clear idea of what the table present. It can be easily interpreted by a common person in less time.
- **To facilitate comparison:** Since the table is divided into rows and columns, for each row and column there is total and subtotal, the relationship between different parts of data can be done easily.
- T**o bring out essential features of data:** It brings out main features of data. It presents facts clearly and precisely without textual explanation.
- **To give identity to the data:** when the data are arranged in a table with title and number, they can be differently identified.
- **To save space:** Table saves space without sacrificing the quality and quantity of data.

Q5. Explain in detail the differnet parts of tables.

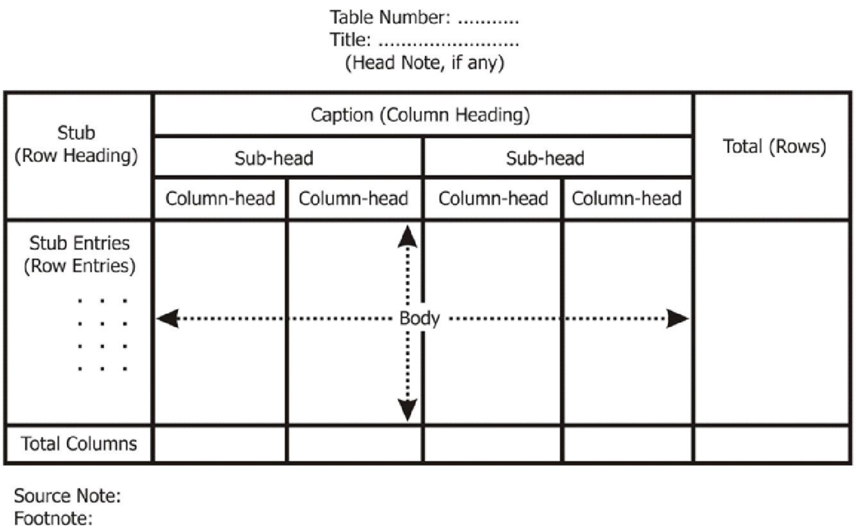
Ans : **(Jan.-20)**

Generally, a table should be comprised of the following components:

1. **Table Number:** Each table must be given a number. Table number helps in distinguishing one table from other tables. Usually tables are numbered according to the or der of their appearance in a chapter. For example, the first table in the first chapter of a book should be given number 1.1 and second table of the same chapter be given 1.2. Table number should be given at its top or towards the left of the table.
2. **Title of the Table:** The title is a description of the contents of the table. Every table must be given suitable title. A complete title has to answer the questions what categories of statistical data are shown, where the data occurred and when the data occurred. The title should be clear and brief. It is placed either just below the table number or at its right.
3. **Caption:** Caption refers to the column headings. It may consist of one or more column headings. Under one column there may be sub heads. The caption should be clearly defined and placed at the middle of the column. If the different columns have different units, the units should be mentioned with the captions.
4. **Stub:** Stub refers to the rows or row heading. They are at extreme left of the table. The stubs are usually wide than column headings but they are as narrow as possible.
5. **Body:** It is most important part of the table. It contains number of cells. Cells are formed by intersection of rows and columns. The body of the table contains numerical information.
6. **Headnote:** It is used to explain certain points relating to the wholetable that have not included in the title, in the caption or stubs. It is placed below the title or at the right hand corner of the table. For example, the unit of measurement is frequently written as a headnote, such as "in thousands", "in crores", etc.

7. **Footnotes:** It helps in clarifying the point which is not clear from the title, captions or stubs. It is placed at the bottom of a table.

There are different ways of identifying the footnotes. One is numbering them consecutively with small numbers 1, 2, 3 or letters a, b, c, d. Another way identifies the first footnote with one star (*), second footnote with two stars (**), third footnote with three stars (***) and so on. Sometimes instead of * , +,@,£ etc used.



3.1.1 Tabulation of Univariate, Bivariate and Multivariate Data

Q6. Explain univariate tabulation with an example.

Ans :

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**

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.

Q7. 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 separate 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.

Q8. Differentiate between univariate and multivariate data.

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

Differences between univariate and bivariate data.

Sl.No.	Univariate Data	Sl.No.	Bivariate Data
1.	It involving a single variable	1.	It involving two variables
2.	It does not deal with causes orrelationships	2.	It deals with causes or relationships
3.	The major purpose of univariateanalysis is to describe	3.	The major purpose of bivariate analysis is to explain
4.	Central tendency - mean, mode, median	4.	Analysis of two variables simultaneously
5.	Dispersion - range, variance, max, min, quartiles, standard deviation.	5.	Correlations comparisons, relationships, causes explanations
6.	Bar graph, histogram, pie chart, line graph, box-and-whisker plot	6.	Tables where one variable is contingent on the values of the other variable.

Q9. What is frequency distribution? Explain the formation of discrete and continuous frequency distribution.

Ans : **(May-22)**

Meaning

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

i) 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	|||	3
1	~~||||~~ ||	7
2	~~||||~~ |||	8
3	~~||||~~ ~~||||~~	10
4	~~||||~~	5
5	~~||||~~	5
6	||	2
Total		40

Table: Discrete Frequency Distribution

ii) 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 deter-mined 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’.

Q10. Explain how ‘ogives’ are drawn for any frequency distribution.

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

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.

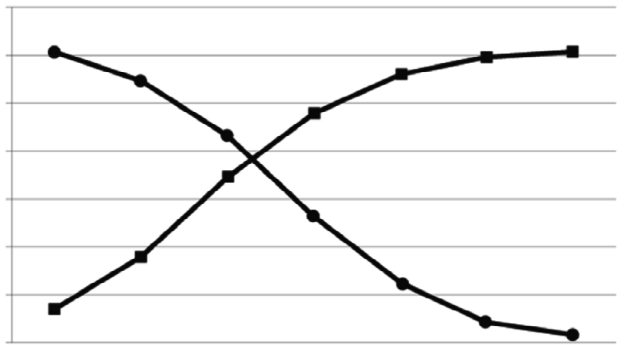
Ogive Graph

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.

PROBLEMS

1. Prepare the Cumulative Frequency.

Table for the following data :

Marks	10-29	30-49	50-69	70-89	90-100
No. of Students	5	25	45	20	5

Sol : **(May-22)**

Marks	No. of students	Cumulative frequency
10 - 29	5	5
30 - 49	25	30
50 - 69	45	75
70 - 89	20	95
90 - 100	5	100

2. For the given data below on the monthly income of 600 families of a rich locality in Hyderabad City, draw :

(i) Less than Ogive and

(ii) More than Ogive Curves and point out the Median.

Monthly Income (Rs. Crore)	No. of Families
Below 10	50
10 - 20	100
20 - 30	200
30 - 40	150
40 - 50	50
50 - 60	30
60 - Above	20

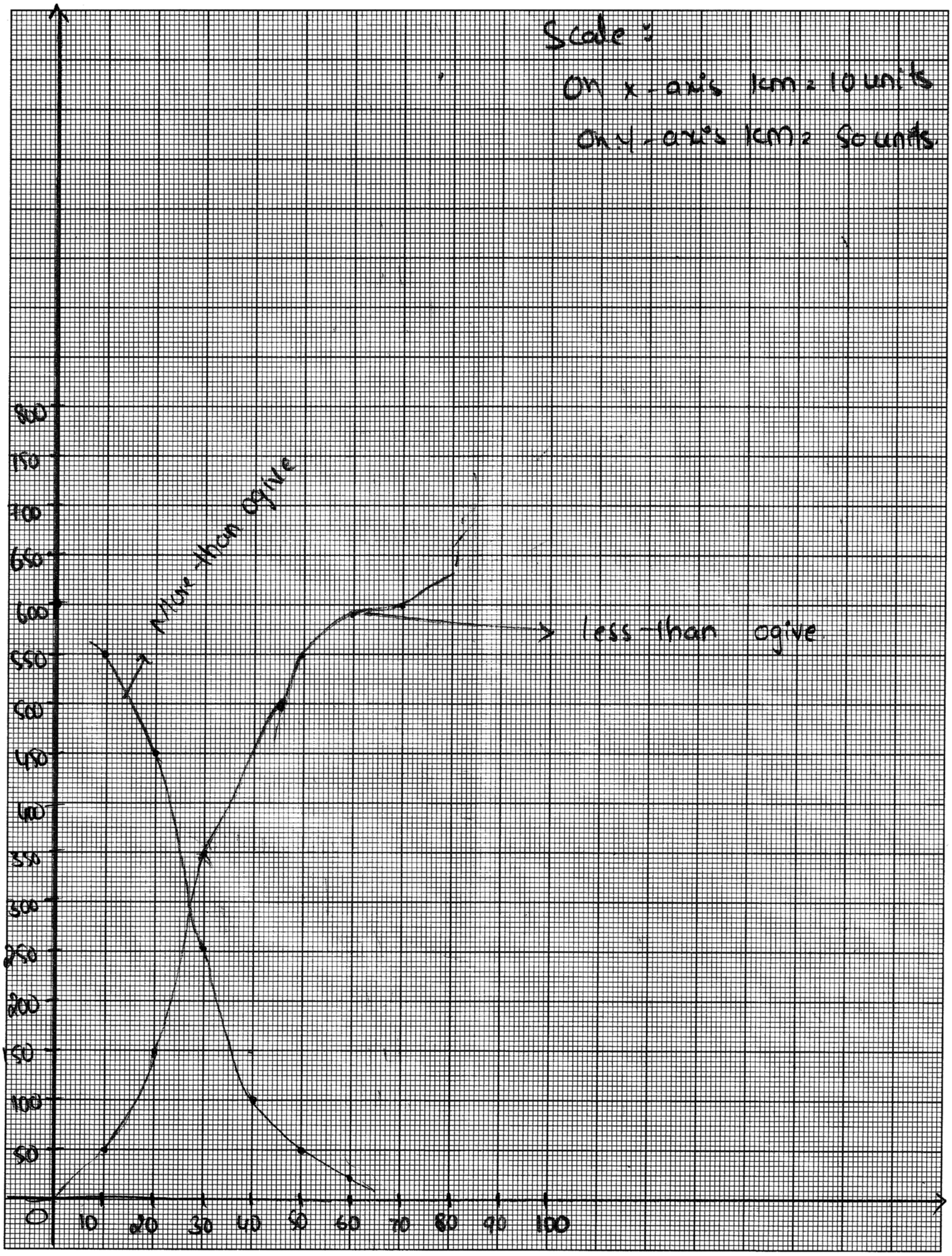
Sol : **(May-22)**

Calculation of lessthan cumulative value

Class interval	Frequency	Cumulative frequency
0 – 10	50	50
10 – 20	100	150
20 – 30	200	350
30 – 40	150	500
40 – 50	50	550
50 – 60	30	580
60 – 70	20	600

Calculation of greaterthan cumulative Value

Class interval	Frequency	Cumulative frequency
0 – 10	50	600
10 – 20	100	550
20 – 30	200	450
30 – 40	150	250
40 – 50	50	100
50 – 60	30	50
60 – 70	20	20


Scale :
On x - axis 1cm = 10 units
On y - axis 1cm = 50 units
More than ogive
less than ogive
0
10
20
30
40
50
60
70
80
90
100
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800

3.2 Diagrammatic and Graphical Representation of Data

Q11. Define diagram and diagramatical representation. 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.

Q12. 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.

Q13. What do you mean by graphical representation of data ? State the rules for graphing. What are the advantages and disadvantages of graphs ?

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

Meaning

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 interpolation 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) Graphs provide restricted information.

Q14. 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 represen- tation 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.

Q15. Differentiate between diagrams and graphs.

Ans :

The differences between Diagrams and Graphs are as follows,

Sl.No.	Diagrams	Sl.No.	Graphs
1.	Diagrams are usually constructed on a plain sheet and they are use ful 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 different types of diagrams.	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 g

3.2.1 One-Dimensional, Two-Dimensional and Three-Dimensional Diagrams and Graphs

Q16. State different types of diagrams.

Ans :

In practice, large variety of diagrams is in use. Diagrams are classified on the basis of their length, width and shape. We will discuss the important types of diagrams which are more frequently used. For sake of application and simplicity several types of diagrams are categories under the following heads.

1. One dimensional diagrams
2. Two dimensional diagrams
3. Three dimensional diagrams
4. Pictogram or Ideographs
5. Cartograms or Statistical Maps

Q17. Explain briefly about one-dimensional diagrams with an examples.

Ans :

This is the most common type of diagrams. They are called one-dimensional diagrams because only length of the bar matters and not the width. For large number of observations lines may be drawn instead of bars to save space.

Merits of Bar diagrams

1. They are easily understood.
2. They are simplest and easiest to make
3. They are simplest and easiest in comparing two or more diagrams.

Types of Bar Diagrams

a) Simple bar diagram
b) Subdivided bar diagram
c) Multiple bar diagram
d) Percentage bar diagram
e) Deviation bars

a) Simple Bar Diagram

A simple bar diagram is used to represent only one variable. It should be kept in mind that, only length is taken into account and not width. Width should be uniform for all bars and the gap between each bar is normally identical. For example the figures of production. Sales, profits etc for various years can be shown by bar diagrams.

Example

Prepare a simple bar diagram for following data related to wheat exports.

Year	Exports (in million tons)
2001	177
2002	219
2003	420
2004	326
2005	202
2006	225

Sol :

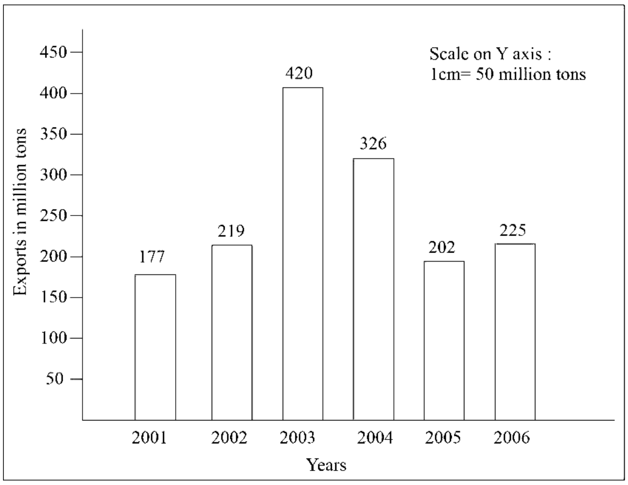


Fig.: Simple Bar Diagram Showing the Wheat Exports in Different Years

b) Subdivided Bar Diagram

In this diagram, one bar is constructed for total value of the different components of the same variable. Further, it is subdivide impropriation to the various components of that variable.

A bar is represented in the order of magnitude from the largest component at the base of the bar to the smallest at the end of the bar, but the order of various components in each bar is kept in the same order. Different shades or colors are used to distinguish between different components. To explain such differences, the index should be used in the bar diagram. The subdivided bar diagrams can be constructed both on horizontal and vertical bases.

Example

The following data shows the production of rice for the period 2010 to 2018. Represent the data by a subdivided bar diagram.

Year	Non-Basmati Rice (in Million metric tons)	Basmati Rice (in Million metric tons)	Total (in Million metric tons)
2010	29	35	64
2011	35	33	68
2012	25	35	60
2013	40	30	70
2014	42	32	74
2015	32	40	72

Sol :

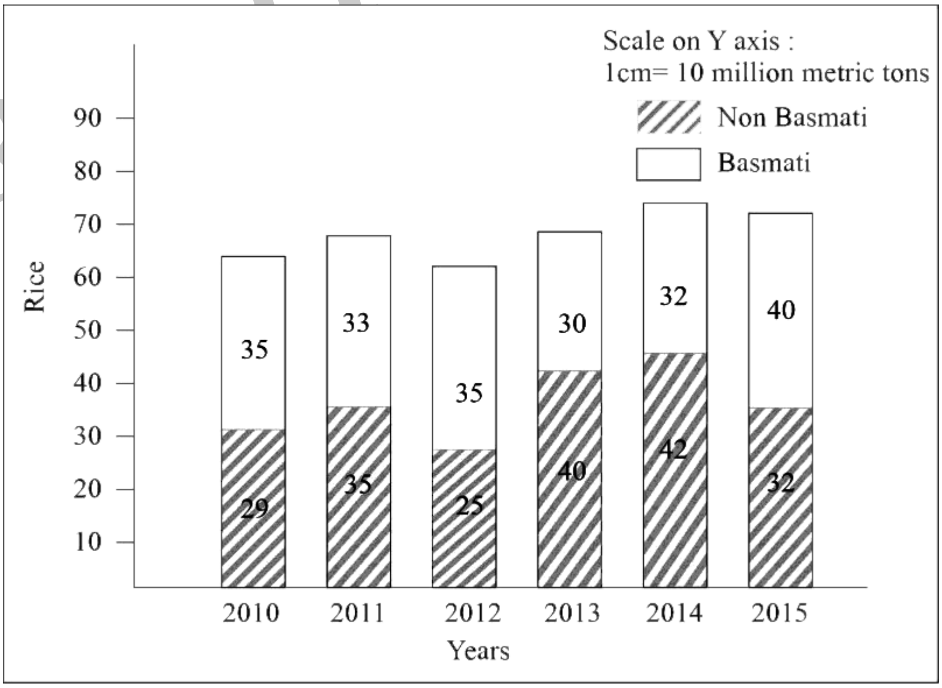


Fig.: Subdivided Bar Diagram Showing the production of Rice (in Different Years)

c) Multiple Bar Diagram

Whenever the comparison between two or more related variables is to be made, multiple bar diagram should be preferred. In multiple bar diagrams two or more groups of interrelated data are presented. The technique of drawing such type of diagrams is the same as that of simple bar diagram. The only difference is that since more than one components are represented in each group, so different shades, colors, dots or crossing are used to distinguish between the bars of the same group.

Example

Represent the following data by a multiple bar diagram.

Class	Physics	Chemistry	Mathematics
Student A	50	63	57
Student B	55	60	68
Student C	48	60	55

Sol :

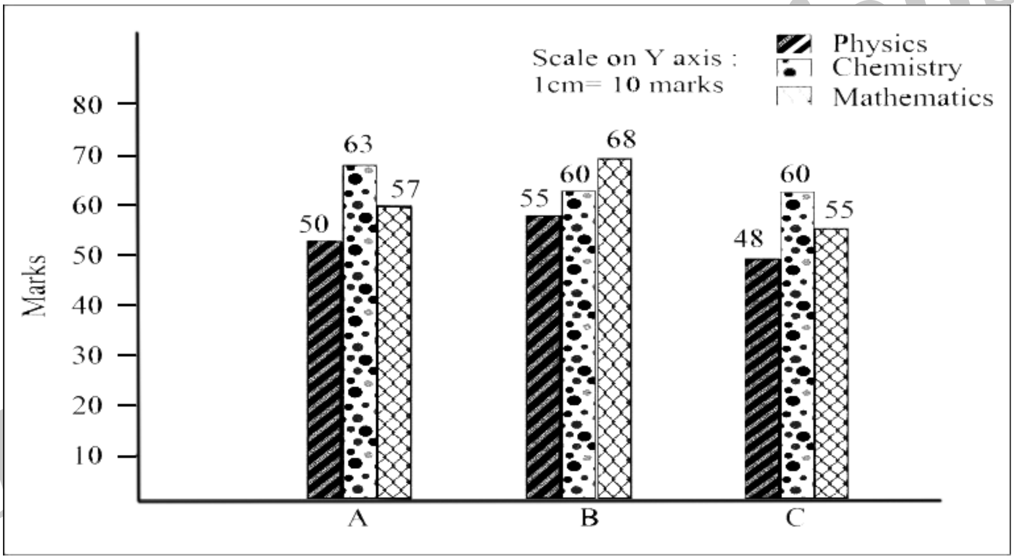


Fig.: Multiple Bar Diagram

d) Percentage Bar Diagram

Percentage bars are particularly useful in statistical work which requires the representation of the 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 percentages of an average.

Example: Draw percentage bar diagram for following data.

Particulars	Cost Per Unit (2010)	Cost Per Unit (2020)
Material	22	35
Lobour	30	40
Delivery	10	20
Total	62	95

Sol :

Express the values in terms of percentage for both the years.

Particulars	Cost Per Unit (2010)	% Cost	Cumulative % cost	Cost Per Unit (2020)	% Cost	Cumulative % cost
Material	22	35.48	35.48	35	36.84	36.84
Lobour	30	48.38	83.86	40	42.10	78.94
Delivery	10	16.12	99.98	20	21.05	99.99
Total	**62**	**100**		**95**	**100**	

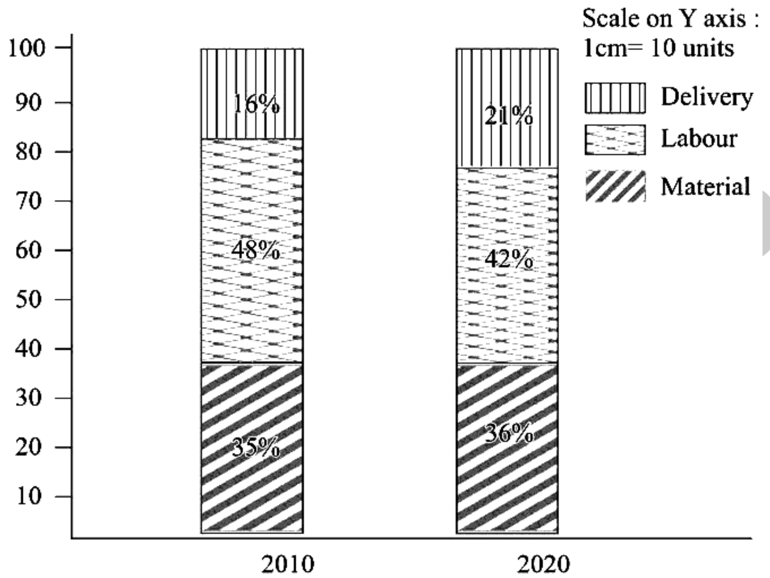


Fig.: Percentage Bar Diagram

e) Deviation Bar Diagram

Deviation bars are used for representing net quantities – excess or deficit. i. e. net profit, net loss, net exports or imports etc For representing net quantities excess or deficit, i.e. net profit, net loss, net exports, net imports, etc., This kind of bars represent both positive and negative values. The values which are positive can be drawn above the base line and negative values can be drawn below it.

Example

Draw deviation bar diagram for following data.

Year	Sales	Profits
2010	24%	29%
2011	15%	-10%
2012	23%	-5%

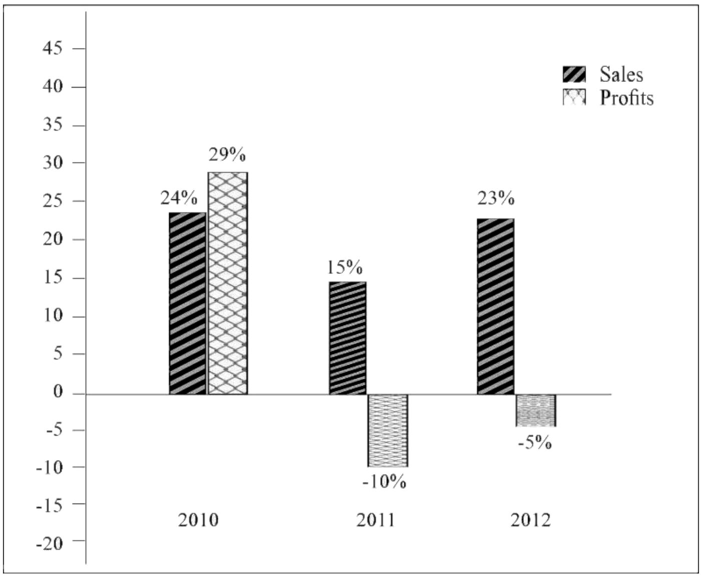


Fig.: Deviation Bar Diagram for sales and profits

Q18. Explain briefly about two-dimensional diagrams with an examples.

Ans :

Meaning

In one-dimensional diagrams, only length of the bar is considered and comparison of bars are done on the basis of length only. In two-dimensional diagrams the length as well as width of the bars is considered. Thus, the area of the bars represent the given data. Two-dimensional diagrams are also known as surface diagrams or area diagrams.

Types of Two Dimensional Diagrams

(a) Rectangles

(b) Squares

(c) Circles

a) Rectangles

In rectangle diagram, given numerical figures are represented by areas of the rectangles. We know that area of rectangle = length x width. While constructing such a diagram both length and width are considered. We may represent the figures as they are given or may convert them to percentage and then subdivide the length into various components.

Example

Represent the following data of monthly expenditure (in rupees) of two families by suitable diagram.

Expenditure	Family A	Family B
Food	2400	1800
Clothing	1600	1200
Education	1000	800
Electricity	200	200
Miscellaneous	800	500

Sol :

First convert the figures into percentage and take the cumulative sum of percentage.

Expenditure	Family A			Family B		
	Rs.	%	Cumulative %	Rs.	%	Cumulative %
Food	2400	40	40	1800	40	40
Clothing	1600	27	67	1200	27	67
Education	1000	17	84	800	18	85
Electricity	200	3	87	200	4	89
Miscellaneous	3200	13	100	500	11	100
Total	**6000**	**100**		**4500**	**100**	

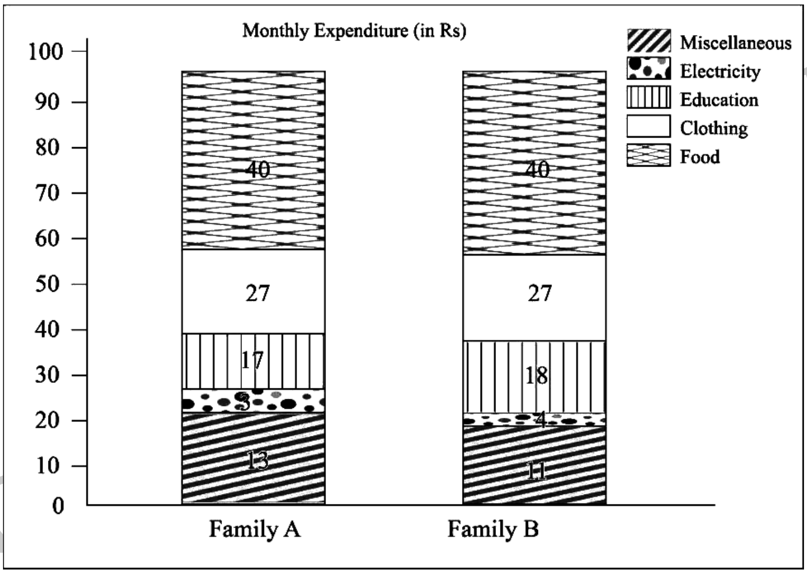


Fig.: Diagram of Rectangles

ii) Squares

The rectangular method of diagrammatic presentation is difficult to use where the values of items vary widely. The method of drawing a square is simple. Take square roots of the given numerical observations as sides of the corresponding squares and then select a suitable scale to draw the squares.

Example

Represent the following data of the number of hospitals in a city in 2000-05, 2005-10, 2010-15 and 2015-20 in square diagram.

Year	No. of Hospitals
2000-05	16
2005-10	64
2010-15	400
2015-20	576

Sol :

Since there is a big gap between first year and last year, a square diagram is suitable here. To decide the side of a square consider following calculations.

Year	No. of Hospitals	Square Root	Side of square in cms = Square Root/4
2000-05	16	4	1
2005-10	64	8	2
2010-15	400	20	5
2015-20	576	24	6

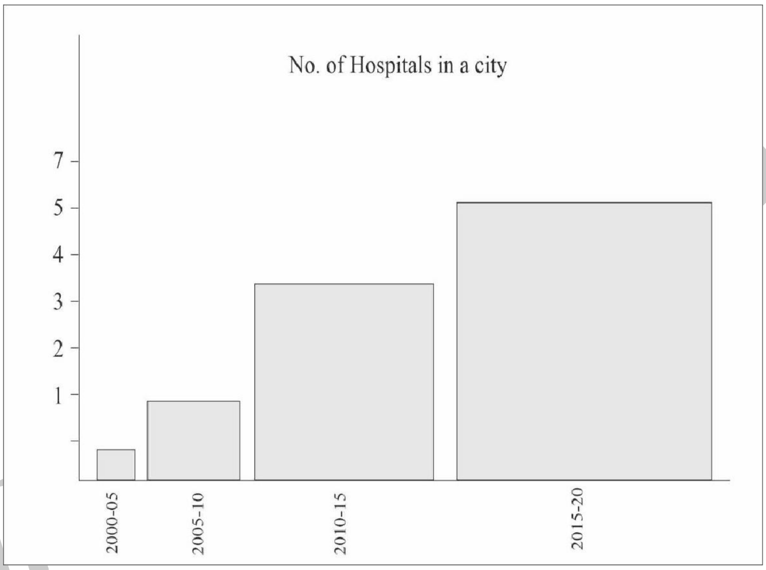


Fig.: Diagram for Squares

C) Circles

As in square diagram, we took given figures/observations as the areas of the corresponding squares. Similarly, here we take given numerical figures/ observations as areas of the corresponding circles. The area of a circle is proportional to the squares of its radius. The radius of circles can be obtained by dividing the value of pie and taking square root. Circles can be used in all those cases in which squares are used.

Circles are difficult to compare and as such are not popular. When it is necessary to use circles, they should be compared on an area basis rather than on diameter basis.

Area of a circle = πr^2 where r is radius of circle

$$\therefore \quad r^2 = \frac{\text{Area}}{\pi}, \pi = \frac{22}{7}$$

$$\therefore \quad r = \frac{\text{Area}}{\pi}$$

Example

Represent the following data of the number of hospitals in a city in 2000-05, 2005-10, 2010-15 and 2015-20 in circle diagram.

Year	No. of hospitals
2000-05	16
2005-10	64
2010-15	400
2015-20	576

Sol.:

Year (I)	No. of Hospitals (n) (II)	$n / (\frac{22}{7})$ (III)	Square Root of $[n / (\frac{22}{7})]$ (IV)	Col (IV)/2 (V)
2000-05	16	5.09	2.25	1.125
2005-10	64	20.36	4.51	2.255
2010-15	400	12	11.28	5.64
2015-20	576	183.27	13.54	6.77

[**Note:** To get smaller value of radius of the circle divide each figure in Col (IV) by 2]

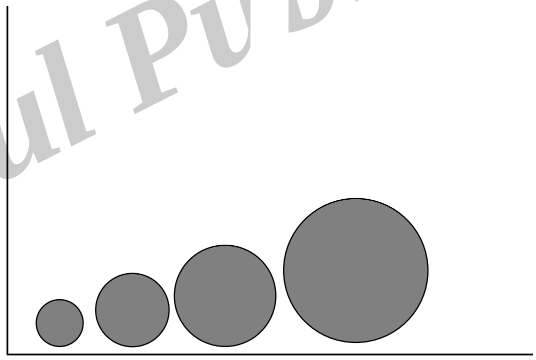


Fig.: for Circles

Q19. Define piediagram. How to create a pie diagram with an example.

Ans :

A pie Diagram is a type of graph that displays data in a circular graph. The pieces of the graph are proportional to the fraction of the whole in each category. Pie diagrams are very popularly used in practice to show percentage breakdown. While making comparisons, pie diagrams should be used on a percentage basis and not on absolute basis.

Steps

1. Take a total of all observations
2. Divide each observation by total and multiply by 100 to get percent. (if instead of percentage, observations are given)

3. Next to know how many degrees for each "pie sector" we need, we will take a full circle of 360° and follow the calculations as below:
4. The central angle of each component = (Value of each component/sum of values of all the components) x 360°
5. Draw a circle of appropriate size with compass and use the protractor to measure the degree of each sector.

In laying out the sectors for pie diagram, it is common practice to begin the largest component sector of pie diagram at 12 O'clock position on the circle. The other component sectors are placed in clockwise direction in descending order of magnitude. Give descriptive label for identification of each sector.

Example

Draw the pie diagram for the following data of cost of construction of house.

Bricks	15%
Steel	35%
Cement	20%
Labour	20%
Supervision	10%

$$\frac{\text{Percentage outlay}}{100} \times 360 = \text{Percentage outlay X3.6}$$

Sector	Percentage	Angle outlay
Bricks	15	15 x 3.6 = 54
Steel	35	35 x 3.6 =126
Cement	20	20 x 3.6 =72
Labour	20	20 x 3.6 =72
Supervision	10	10 x 3.6 =36
Total	**100**	**360**

(Note: The angles have been arranged in ascending order)

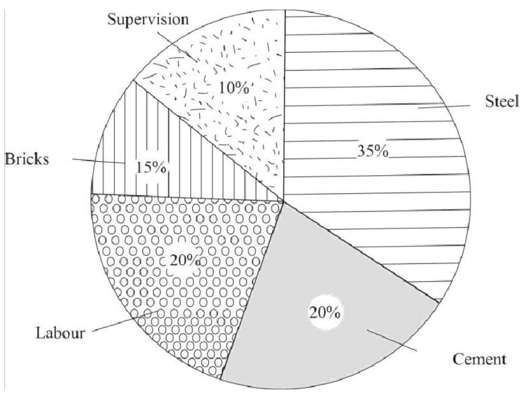


Fig.: Pie Diagram

Q20. Explain briefly about three-dimensional diagrams with an examples.

Ans :

In three dimensional diagrams three things namely, length, width and height have to be considered. Three-dimensional diagrams, also known as volume diagrams consist of cubes, cylinders, spheres etc. Such diagrams are used where the range of difference between the smallest and the largest value is very large. For example, if two values are in the ratio of 1:1000 and if bar diagram are used to represent them, the shortest bar would be of one-thousandth part of the largest bar. If squares or circles are used then the side of the square or the radius of one circle would be proportionately too large or too small than the other. If cubes are used then their sides would be in the ratio of 1:10. This example makes it clear that three-dimensional diagrams have an important role to play when the gap between the smallest and the largest value is very large.

The disadvantage of three-dimensional data is the side of a cube must be proportionate to the cube root of the magnitude to be represented. It is very difficult for the eye to read precisely such diagrams and hence they are not recommended for statistical presentation.

PROBLEMS

3. 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 :

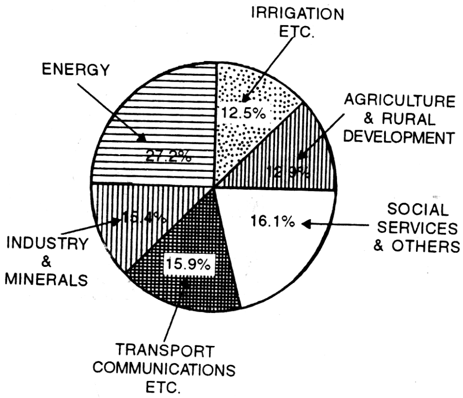
The angle at the centre is given by

$$\frac{\text{Percentage outlay}}{100} \times 360 = \text{Percentage outlay} \times 3.6$$

Computation for PIE Diagram

Sector	**Percentage**	**Angle outlay**
Agriculture and Rural Development	12.9	$12.9 \times 3.6 = 46^{o}$
Irrigation, etc.	12.5	$12.5 \times 3.6 = 45^{o}$
Energy	27.2	$27.2 \times 3.6 = 98^{o}$
Industry and Minerals	15.4	$15.4 \times 3.6 = 56^{o}$
Transport, Communication, etc.	15.9	$15.9 \times 3.6 = 57^{o}$
Social Services and others	16.1	$16.1 \times 3.6 = 58^{o}$
Total	**100.0**	**360^{o}**

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).

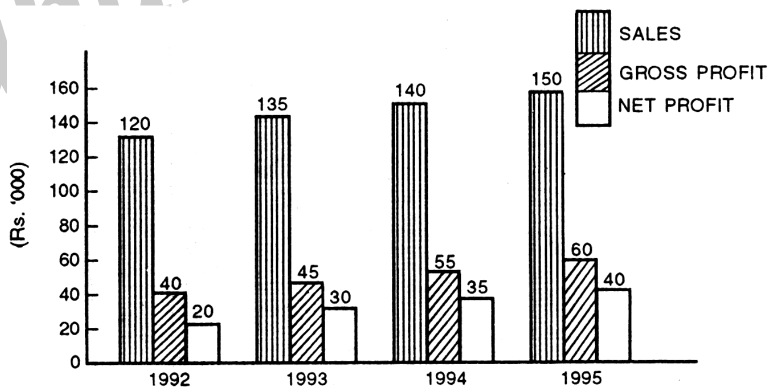


4. Draw a multiple bar diagram from the following data :

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)



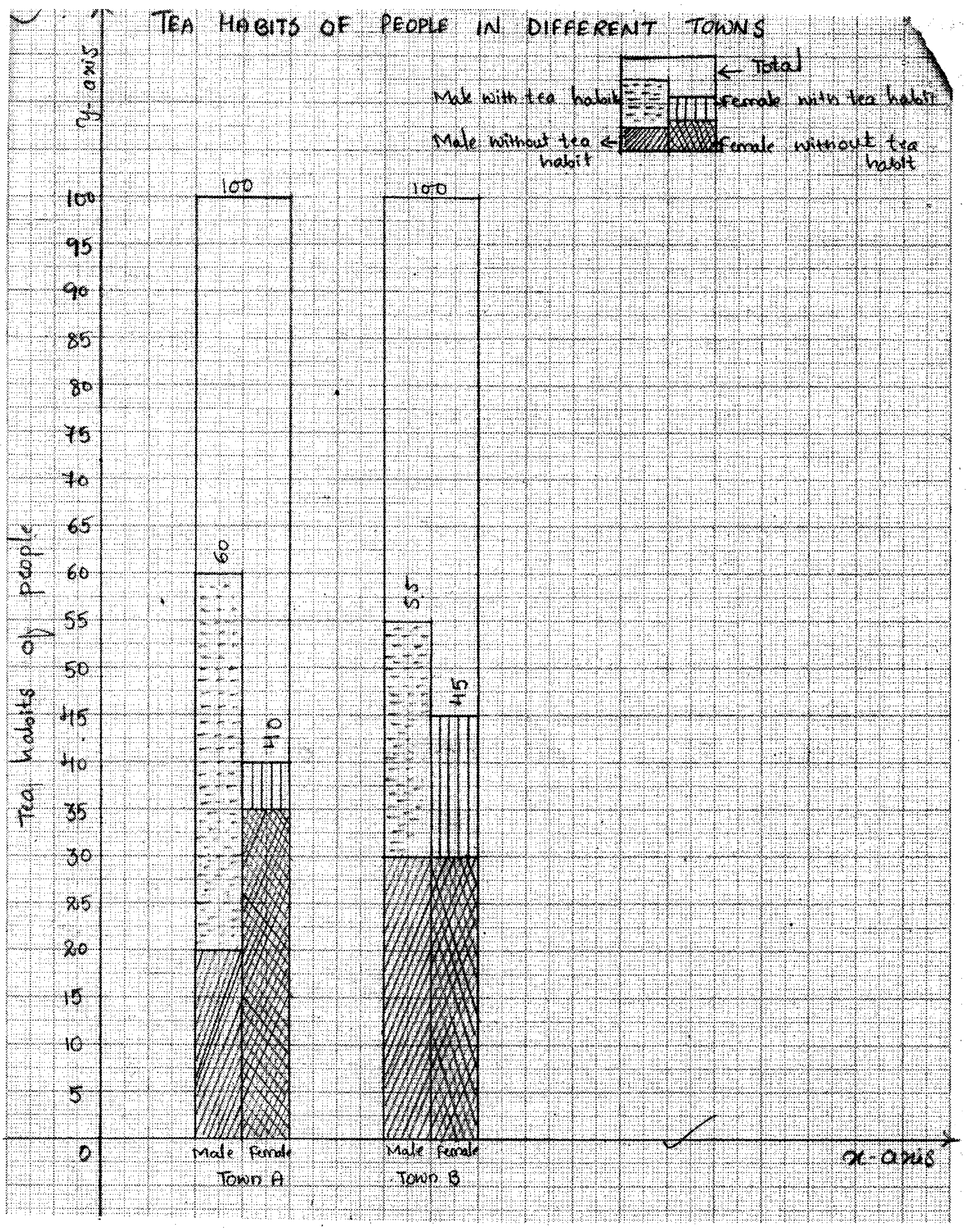
5. Represent the data shown in the following table by using suitable graphical method.

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

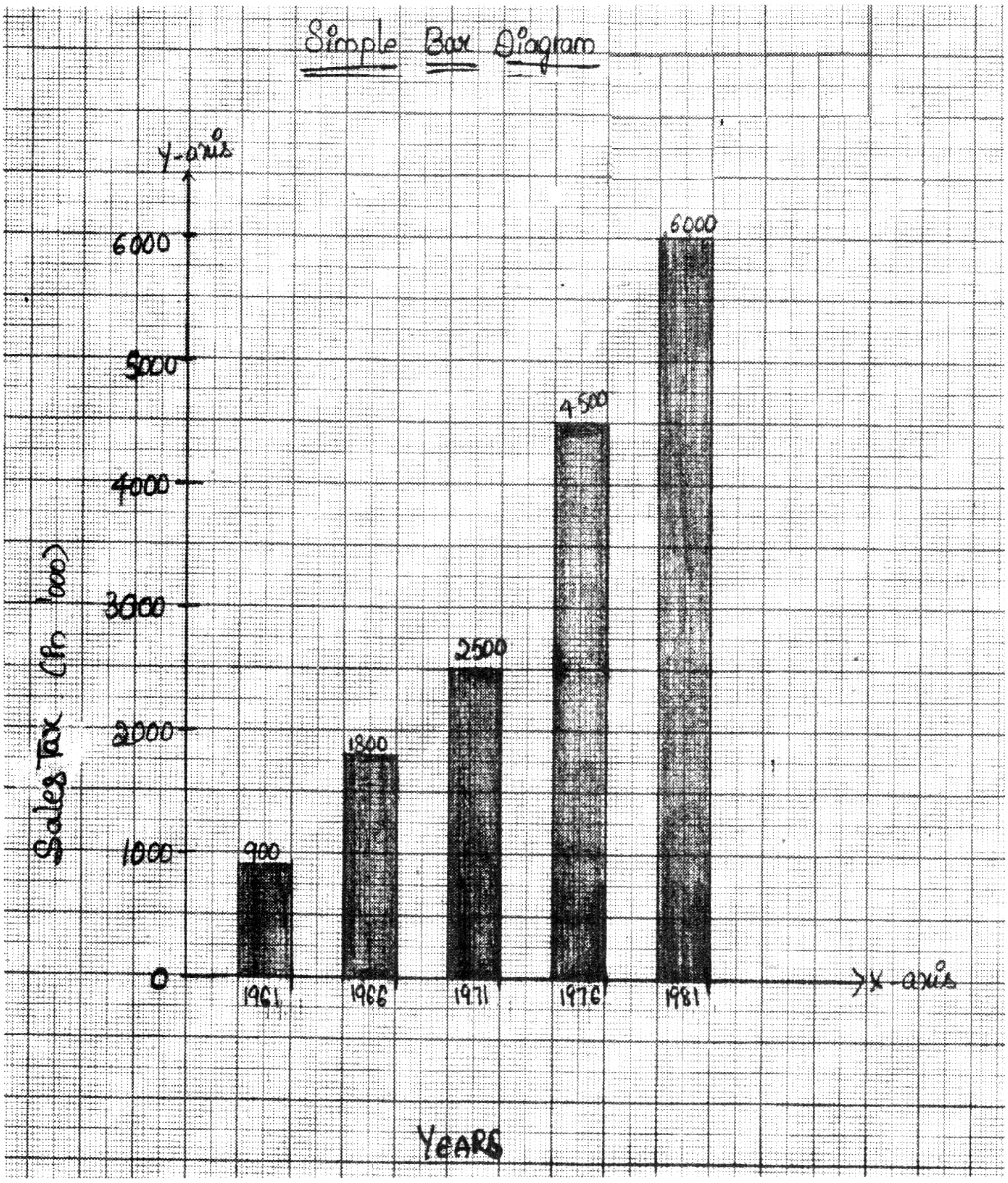
Department	Town A			Town B		
	Male	Female	Total	Male	Female	Total
No. of people with tea habit	40	5	45	25	15	40
No. of people with - out tea habit	20	35	55	30	30	60
Total	**60**	**40**	**100**	**55**	**45**	**100**



6. 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 :



7. Draw pie diagram to represent the following data of proposed expenditure by a State Government for the year 1999 - 2000.

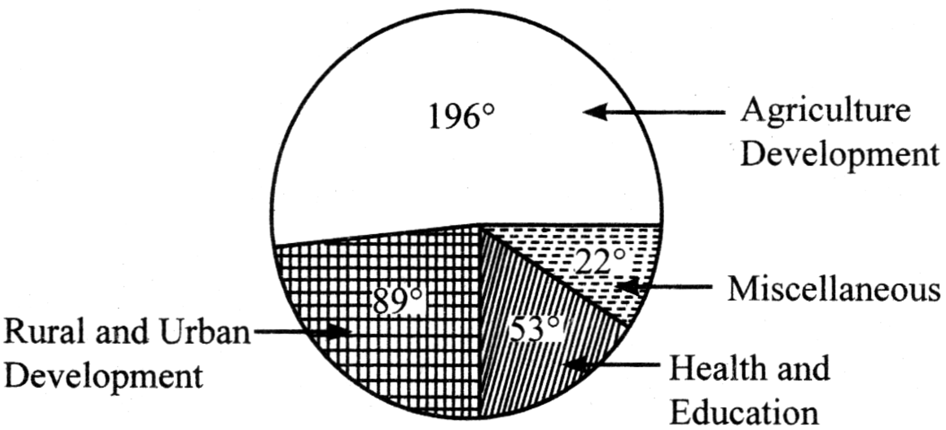
Items	Agriculture Development	Rural and Urban Development	Health and Education	Miscellaneous
Proposed Expenditure (in Million Rupees)	4400	2000	1200	500

Sol :

Calculations for Pie Chart

Items	Proposed Expenditure (in Million Rupees)	Angle at the Centre
(a)	(b)	(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°**

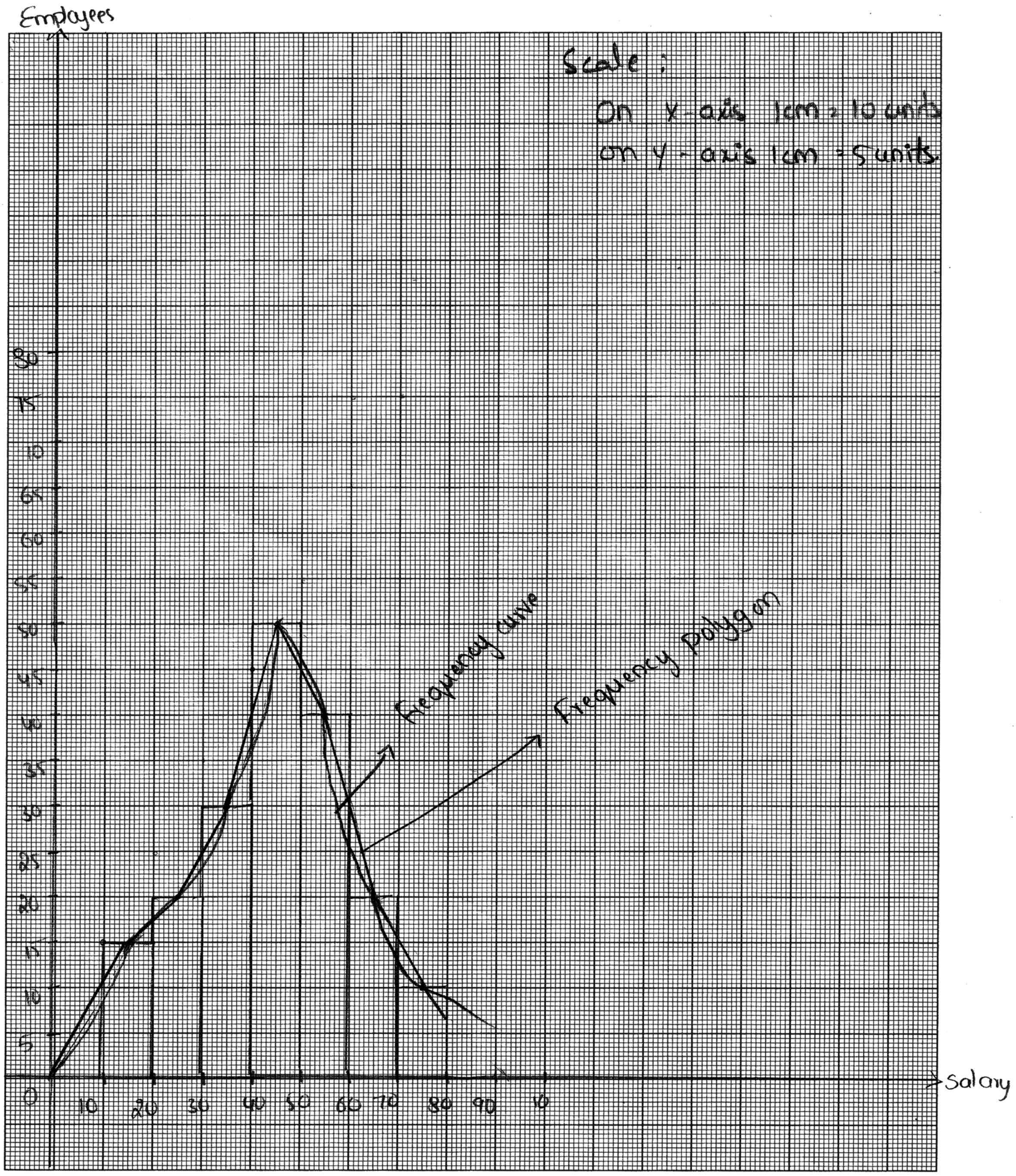
Pie diagram representing proposed expenditure by state government on different items for 1999 - 2000.



8. For the following data, draw Histogram, Frequency Curve and frequency Polygon.

Salary (Rs. in '000)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Employees	15	20	30	50	40	20	10

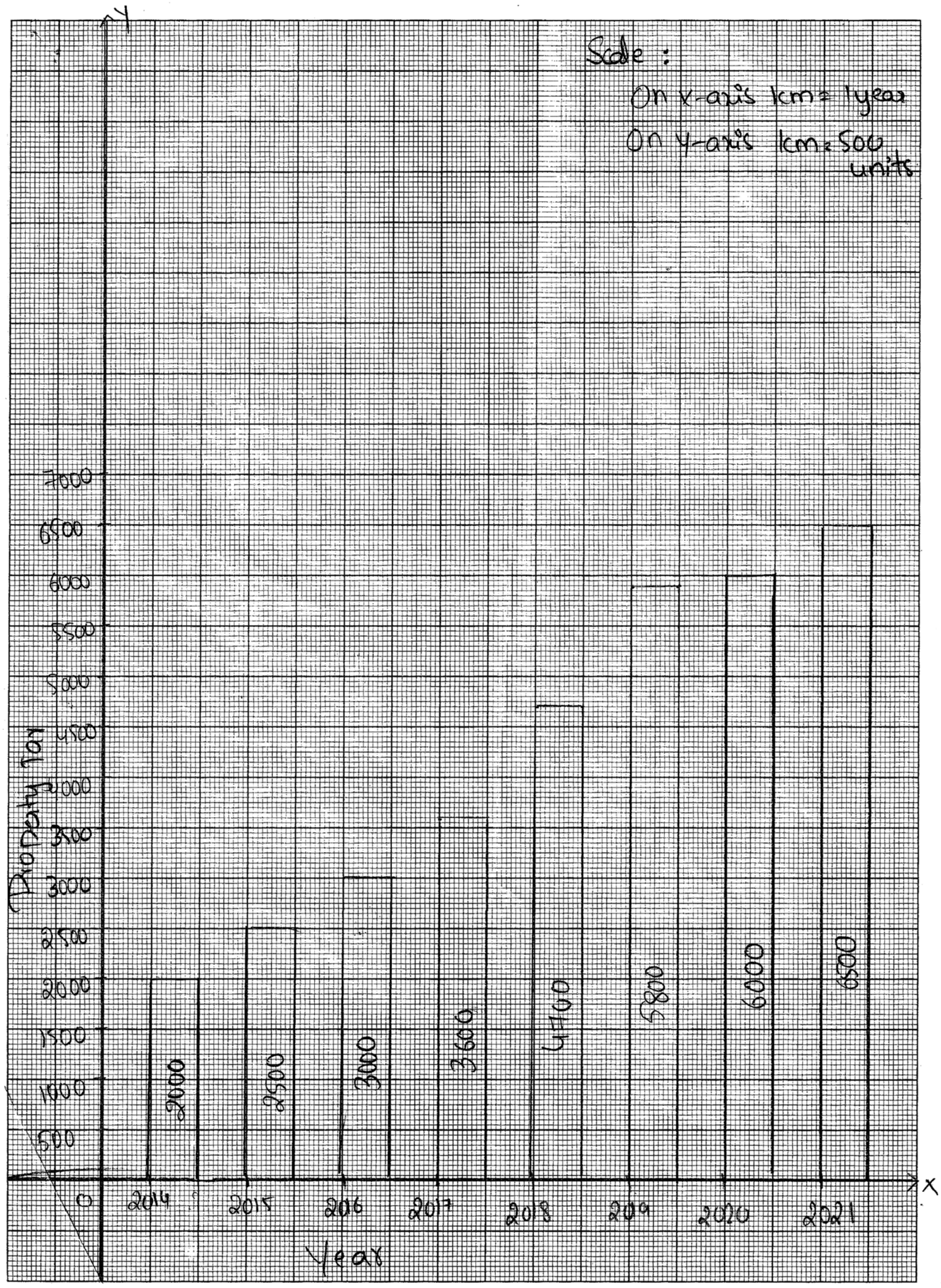
Sol : **(May-22)**



9. For the following data, prepare a Bar Diagram

Year	2014	2015	2016	2017	2018	2019	2020	2021
Property Tax (Rs. Crore)	2000	2500	3000	3600	4700	5800	6000	6500

Sol : **(May-22)**



3.3 Introduction to Statistics

Q21. Define Statistics. Explain the characteristics of Statistics.

Ans :

Introduction

The. term 'Statistics' has been defined differently by different authors. Some have defined the word in the sense of numerical data, whereas others have defined in the sense of statistical methods.

Definitions

A few definitions have been discussed here.

i) **According to Bowel defined Statistics** as 'Numerical statement of facts in any department of enquiry placed in relation to each other" According to Yule and Kendall "By statistics we mean quantitative data affected to a marked extent by multiplicity of causes'.

ii) A comprehensive definition is given by Prof. Horace Secrets that points out all characteristics that numerical data must possess to be called statistics.

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

Characteristics

1. **Statistics are aggregates of facts** : single and unconnected figures are not statistics, for example, if it is stated that Raj, a student, secured 50 marks, it is not statistics. However, marks secured by the students of the class would constitute statistics. A single figure relating to production income, marks height, etc. cannot be regarded as statistics but aggregates of such facts would be regarded as statistics.

2. **Statistics are affected to a marked extent by multiplicity of causes** : Numerical facts are affected by a multiplicity of factors. For example, the price, of a commodity is affected by number of factory such as supply, demand, imports, exports, money in circulation, competitive products in the market and so on. It is very difficult to study separately the effect of these factors on the price of the commodity. In physical sciences, it is possible to isolate the effect of various factors on a single item, but statistics are commonly used in social sciences and in social sciences it is very difficult to study the effect of any one factor separately. In statistical methods, the effects of various factors affecting a particular phenomenon are generally studied in a combined form, though attempts are also made to study the effects of different factors separately as well.

3. **Statistics are numerically expressed :** Only numerical data constitutes statistics. Qualitative expressions like good, bad, young, old etc. Cannot be regarded as statistics. Statement like the standard of living of people in India has improved 'or production of petroleum products has increased in India do not constitute statistics. But Statements like "Rice production in India in 2003-04 is expected to be 86.4 million tones as against 72.66 million tons in 2002-03" (Source Economic Survey 2003-2004) is statistics.

4. **Statistics are enumerated or estimated according to reasonable standards of Accuracy:** Numerical information can be either enumerated or estimated. If they are enumerated i.e., actually counted or measured, they are supposed to be exact and accurate. If complete enumeration is not possible because of the large size of high cost, then data is estimated by using sampling technique. Estimated figures cannot be absolutely accurate and precise. The degree of accuracy depends to a large extent on the particular purpose for which are information is collected and the nature of the particular problem about which the data is collected. There cannot be uniform standard of accuracy for all types of data collection. For example, while calculating the marks of students in an entrance examination or the number of votes received by a candidate in an election, it is necessary that it should be absolutely accurate, but while calculating the number or persons watching a cricket match on Television, the numbers need not be accurate.

5. **Statistics are collected in a systematic manner :** The data should be collected in a very systematic manner. For any socioeconomic survey, a proper schedule depending on the object of enquiry should be prepared and trained investigators should be used to collect the data. An attempt should be made to reduce personal bias to the minimum. Data collected in haphazard manner may give inaccurate results.

6. **Statistics are collected for a pre-determined purpose :** The Purpose of collecting the date should be pre-determined. Otherwise, the data collected may not serve any purpose and may become useless. One should not waste time and money in collecting information that is irrelevant for the enquiry. For example, if the purpose of enquiry is to measure the cost of living of low-income group people, we should collect information about the items that are generally consumed by them. Hence for such an index it is useless to collect information on items such as cars, refrigerators, cell-phones etc.

7. **Statistics should be placed in relation to each other** : Statistical information is collected mostly for the purpose of comparison. If the data collected cannot be compared, then much of the purpose of collection will be lost. The information collected should be homogeneous and not heterogeneous in character. Statistical data are often compared period wise or region-wise. For example, the data relating to population of a country for different years or population of different countries in some fixed period will be regarded as statistics. But data relating to the sue of shoe of an individual and his intelligence quotient do not constitute statistics.

Q22. Explain the scope of Statistics.

Ans :

There is hardly and field that has remained untouched by Statistics. Statistics is viewed not just as a device for collecting numerical data, but as a means of sound techniques for their handling, analysis and for drawing valid inference from them. From this perspective, the scope or subject matter to statistics can be broadly studied under 2 hands namely.

(A) Statistical Methods and

(B) Applied Statistics

(A) Statistical Methods

Statistical methods are the tools that are in the hands of the statistician. They include all the general principles and techniques that are commonly used in the collection, analysis and interpretation of data. These methods are applicable to all kinds of data. The stage involved in study of any kind of data are:

(i) Observation and Collection

(ii) Organization

(iii) Presentation

(iv) Analysis and

(v) Interpretation.

These have been explained under the topic 'Definition of Statistics as Statistical Methods.

(B) Applied Statistics

Applied Statistics deals with application of Statistical methods of specific problems or concrete forms. To illustrate, if a software services firm is experiencing attrition (loss of personnel), it may be worthwhile to investigate the reasons for the same. Special techniques can be employed to understand the underlying trends. For example, a correlation analysis could probably indicate that increasing opportunities (growth in the industry) and attrition are highly correlated. All such techniques .and the results obtained by employing such techniques form part of applied Statistics.

Q23. Explain the nature of Statistics.

Ans :

i) Statistics as Science

'Science' is a systematic study of knowledge. It studies cause and effect relationships between different variables and comes up with principles that are universally applicable. The law of gravity, Archimedis principle etc are all examples of the various studies in science. From this stand point, statistics can also be considered as 'Science' Statistics is also involved in study of cause and effect. It has come up with its own set of 'principles' that are widely used across various fields. However, the principles of Science have a very high degree of precision. This is on account of our ability to isolate the factors being studied in a laboratory environment. This is not possible in case of Statistics. Hence, statistical laws or principles cannot be expected to have the same degree of precision. Thus Statistics can be termed as a ' Social Science'. It is more of a 'Scientific Method' Than a science in itself.

ii) Statistics as Art

Art is an expression of skill or creativity. It is not concerned with cause and effect analysis. It is concerned with doing things that cannot be done by everybody. It demands rigorous practice to continuously hone such skills. Painting, Music, story telling etc are examples of Art. It is very evident that a person with many years of rigorous practice would be a better artist than a novice, even if the novices is 'gifted'. Statistics can very well be considered an Art in as much as it is concerned with the skill of handling facts and figures. Statistics are being very innovatively and creatively used in analysis and presentation of data. However, Art is an end in itself. Pursuit of art gives tremendous satisfaction. Statistics is only a means to an end. It is used to arrive at conclusions.

Thus, Statistics is both a Science and an Art. However, it is neither pure science nor pure Art. It is a scientific method that can be artistically used for the benefit of mankind.

Q24. Explain the functions of statistics.

Ans : **(Imp.)**

Following are the important functions of Statistics

1. **Systematic Collection and Presentation of Facts** : Facts backed by objective numbers carry lot more conviction. Statistics ensures that data is' collected in a systematic manner and presented without any subjective bias. Statistics adds a bit of precision and definiteness to general statements, resulting in greater conviction.

2. **Simplification of Mass Figures** : Statistics cuts through the clutter of data normally available and presents the facts effectively. The large mass of data collected is transferred into a few critical figures, which help in overall analysis and interpretation.

3. **Facilitates Comparison :** Statistics enables impartial and fair comparison of data across a wide range of competing alternatives. For example, Inflation rate compares the current prices with prices of previous year and indicates the trend of rising prices. Thus, Statistics performs the function of providing a definite meaning to the state of affairs by quantification of data and comparison.

4. **Helps in Formulation and Testing of Hypothesis :** Statistics has developed into a separate discipline, with its own theories, principles and methodology, which can be used to answer specific queries. For example, if a contractor has to decide between two alternative choices of cement, he has to check whether the quality of cement is satisfactory. He can take the help of statistics and use sampling techniques and principles to arrive at a decision. Similarly, Statistics can be used in determining the result of a medicine on a patient, the impact of newly introduced Government policy on the general public, and whether the purchase of a computer has resulted in efficiency gains etc.

5. **Helps in prediction :** Statistical methods are a very useful tool in attempting to forecast the future. If an investor wishes to purchase the shares of say. Reliance Industries Limited, he can use Statistics to find out whether the share price of Reliance is likely to increase or decrease. He can use correlation and Regression to know the extent of impact of some key factors on the performance of the company and then take his decision. Thus, Statistics is the closest substitute of Crystal Ball.

6. **Helps in Policy Formulation and Decision -Making :** Many a time, Government policy and even business decisions are taken intuitively or on "gut feel" However, Statistics can reveal as to whether such intutive feeling can be translated into reality. For example, Statistics would have revealed that none of the previous Governments of Andhra Pradesh could successfully implement total prohibition of liquor in the state. Similary, a businessman would be better off if he conducts a market research before entering into a particular business.

7. **Enlarge Individual Experience :** A Proper function of Statistics is to enlarge human experience and knowledge. Statistics makes it easier for man to understand, describe and measure the impact of any action. The efficient and sound techniques of Statistics here ensured that different fields of knowledge are opened up to mankind.

8. **Study Relationships between Various Phenomena :** Statistics enables us to observe and understand the relationships between different phenomena. For example, there is a strong relationship between unemployment related data and data pertaining to GDP.

9. **Measures Uncertainty :** Statistics helps in ascertaining the chance of occurrence of and event. It also helps in finding out the impact of the happening or not happening of an event. It helps mankind to cope with uncertainty. Thus, statistics performs the function of systematically creating knowledge through analysis of facts and help the user of statistics in making informed, intelligent decisions.

Q25. Explain the limitations of statistics.

Ans :

The following are some of the limitations.

1. **Statistics does not study qualitative phenomenon :** Statistics are numerical statements of facts. It can be applied only to such problems that can be measured quantitatively. Statistics cannot be used directly for the study of qualitative characteristics such as honesty, beauty, intelligence, culture etc. However, it may be possible to analyze such characteristics indirectly. For example we may study the intelligence of students on the basis of marks secured by them in an examination.

2. **Statistics does not study individual measurements :** A single o.r isolated figure cannot be regarded as statistics unless it is part of the aggregate of fact relating to a field of enquiry. Statistical methods do not give any recognition to an object, person or an event in isolation. The average income of a group of persons might have remained the same over two periods, yet some persons in the group might have become poorer than what they were before, statistical methods ignore such individual cases.

3. **Statistical laws are true only on average:** : Statistics, as a science is not as accurate as many other sciences. Statistical laws are not universally true like laws of physics etc. They are true only on an average. Statistics deals with such phenomena that are affected by a multiplicity of causes and it is difficult to study the effects of each of these factors separately. Due to this limitation, the conclusions arrived at are not perfectly accurate.

The following are some of the limitations.

1. **Statistics does - not study qualitative phenomenon** : Statistics are numerical statements of facts. It can be applied only to such problems that can be measured quantitatively. Statistics cannot be used directly for the study of qualitative characteristics such as honesty, beauty, intelligence, culture etc. However, it may be possible to analyze such characteristics indirectly. For example we may study the intelligence of students on the basis of marks secured by them in an examination.

2. **Statistics does not study individual measurements** : A single o.r isolated figure cannot be regarded as statistics unless it is part of the aggregate of fact relating to a field of enquiry. Statistical methods do not give any

recognition to an object, person or an event in isolation. The average income of a group of persons might have remained the same over two periods, yet some persons in the group might have become poorer than what they were before, statistical methods ignore such individual cases.

3. **Statistical laws are true only on average** : Statistics, as a science is not as accurate as many other sciences. Statistical laws are not universally true like laws of physics etc. They are true only on an average. Statistics deals with such phenomena that are affected by a multiplicity of causes and it is difficult to study the effects of each of these factors separately. Due to this limitation, the conclusions arrived at are not perfectly accurate.

3.4 Measurement of Central Tendency

Q26. Define Average. What are the functions of an average.

Ans : **(Imp.)**

Introduction

Summarisation of the data is a necessary function of any statistical analysis. As a first step in this direction, the huge mass of unwieldy data are summarised in the form of tables and frequency distributions. In order to bring the characteristics of the data into sharp focus, these tables and frequency distributions need to be summarised further. A measure of central tendency (or) an average is very essential and an important summary measure in any statistical analysis. An average is a single value which can be taken as representative of the whole distribution.

Definitions :

The average of a distribution has been defined in various ways. Some of the important definitions are :

(i) **According to Clark and Sekkade,** "An average is an attempt to find one single figure to describe the whole of figures".

(ii) **According to Murry R. Spiegal,** "Average is a value which is typical or representative of a set of data".

(iii) **According to Croxton and Cowden,** "An average is a single value within the range of the data that is used to represent all the values in the series. Since an average is somewhere within the range of data it is sometimes called a measure of central value".

(iv) **According to Sipson and Kafka,** "A measure of central tendency is a typical value around which other figures congregate".

Functions

1. **To present huge mass of data in a summarised from**

It is very difficult for human mind to grasp a large body of numerical figures. A measure of average is used to summarise such data into a single figure which makes it easier to understand and remember.

2. **To facilitate comparison**

Different sets of data can be compared by comparing their averages. For example, the level of wages of workers in two factories can be compared by mean (or average) wages of workers belonging to each of them.

3. **To help in decision-making**

Most of the decisions to be taken in research, planning, etc., are based on the average value of certain variables. For example, if the average monthly sales of a company are falling, the sales manager may have to take certain decisions to improve it.

Q27. Explain the significance of an average.

Ans :

- The measures of central tendency provides a single value which represents the features of the whole group.
- The single/central value is calculated by contracting the mass of data into one single value.
- It helps us to understand briefly about the entire group through one single value.
- The measures of central tendency also provides a comparison which can be carried out at a point of time or over a period of time.

- The comparison may be made between two or more data sets related to each other.
- For example, the percentage profit of different firms producing pens. But, at the time of drawing presumptions from the data, the factors influencing data at different points should also be considered and it should also be compared with the same value of measure of central tendency.
- One measure of central tendency cannot be compared with the different measure of central tendency i.e., one set of values of arithmetic mean cannot be compared with the other set of values of median or mode.
- It implies that a data set should contain only one value of central tendency.
- Every measure of central tendency has its own importance and process for calculation and application.

Q28. What are the characteristics of a good average ?

Ans :

1. It should be rigidly defined

An average should be rigidly defined so that there is no scope for confusion or manipulation. The average value will become very unstable and non representative of the base data if it is not well defined. It is ideal to use an average that is mathematically defined by way of formula.

2. It should be easy to calculate and simple to follow

Calculation of an average should be simple to understand. It should be easy to calculate, preferably without the help of calculators. If an average is too complex to understand and calculate, its use will be very limited. It should also be capable of expression in simple numerical terms without advanced mathematical intricacies.

3. It should be based on all observations in the series

An average will be truly representative of the whole mass of data if it is computed from all the observations.

4. It should not be affected much by an extreme values

An average will be representative of the data only if it can set off extreme values against each other. A few very small or very large observations should not unduly affect the value of a good average.

5. It should be capable of further statistical processing

An average should be capable of being used in calculation of other statistical measures such as standard deviation, correlation, etc.

6. It should be capable of further algebraic treatment

An average should lend itself readily to further algebraic treatment. For example, if averages of two (or) more sets of data are known, it should be possible to obtain the average of combined group.

7. It should possess sampling stability

An average should be least affected by sampling. If we take independent random samples of the same size from a given population and compute the average for each of these samples, the values so obtained from different samples should not be very different from one another.

8. It should be representative of the data

A good average should represent maximum characteristics of the underlying data.

Q29. What are the limitations of averages?

Ans :

1. Misleading Conclusions

Average is a single numerical figure representing the characteristics of a given distribution. This number is vulnerable to errors in interpretation and can lead to misleading conclusions. For example, if the maximum and minimum temperatures of a particular city are 48° F and 2° F respectively, the average may still work out to 24° F. Looking at the average, the weather conditions might look very comfortable, but that is not really the case.

2. Choice of Average

There are different types of averages. Different types of averages are suitable for different objectives. The utility of average depends on a proper and judicious choice of the average. A wrong choice of the average might lead to erroneous results.

3. Incomplete Picture

An average does not provide the complete picture of a distribution. There may be number of distribution having the same average but differing widely in their structure and constitution. To form a complete idea about the distribution, the measures of central tendency are to be supplemented by some more measures such as dispersion, skewness and kurtosis.

4. Inadequate Representation of Data: In certain types of distribution like U- shaped or J-shaped distribution, an average, which is only a single point of concentration, does not adequately represent the centre series.

5. Absurd Conclusions

Sometimes, an average might throw up very absurd results that are not realistic. For example, the average size of a family might work out to a fractional number, which is not realistic.

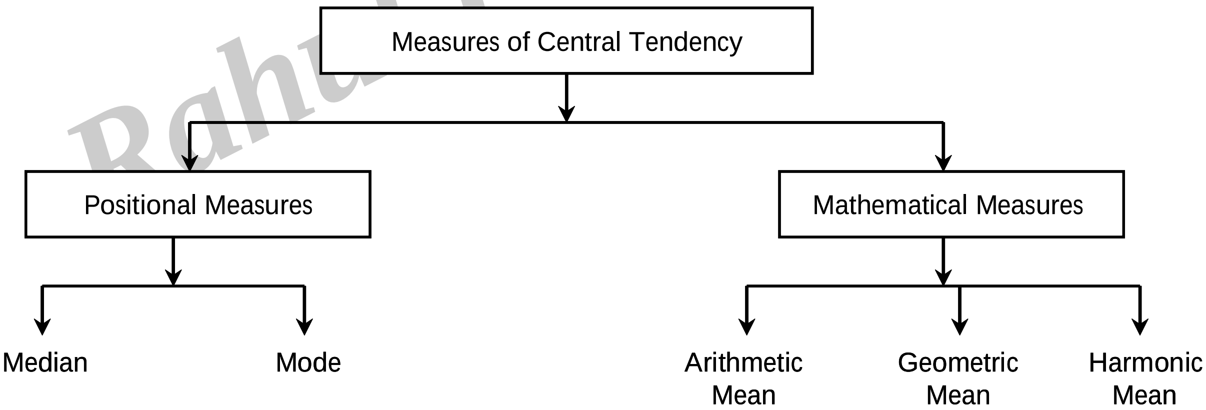
Q30. Explain the different types of averages.

(OR)

What are the various measures of central tendency?

Ans :

Various measures of averages can be classified into the following two categories :



3.4.1 Arithmetic Mean

Q31. What is arithmetic mean? State the merits and demerits of arithmetic mean.

Ans :

Meaning

Arithmetic Average (or) Mean of a series is the figure obtained by dividing the total "Values of the various items by their number. In other words it is the sum of the values divided by their number. Arithmetic means is the most widely used measure of central tendency.

Merits

1. It is rigidly defined. Hence, different interpretations by different persons are not possible.
2. It is easy to understand and easy to calculate. In most of the series it is determinate and its value is definite.
3. It takes all values into consideration. Thus, it is more representative.
4. It can be subjected to further mathematical treatment. The properties of Arithmetic mean are separately explained elsewhere in the chapter.
5. It is used in the computation of various other statistical measures.
6. It is possible to calculate arithmetic average even if some of the details of the data are lacking. For example, it can be known even when only the number of items and their aggregate value are known, and details of various items are not available. Similarly, the aggregate value of items can be calculated if the number of items and the average are known.
7. Of all averages, arithmetic average is least affected by fluctuations of sampling. Thus, it is the most stable measure of central tendency.
8. It provides a good basis for comparison.
9. Arithmetic mean is impacted by every observation. It gives weight to all items in direct proportion to their size.

Demerits

While Arithmetic mean satisfies most of the conditions of an ideal average, it suffers from certain drawbacks. Some of the demerits or limitations of Arithmetic Mean are listed below:

1. It cannot be determined by inspection.
2. It cannot be located graphically.
3. It cannot be used in the study of qualitative phenomena.
4. It can be significant impacted by extreme values and may lead to erroneous conclusions. Abnormal items may considerably affect this average, particularly when the number of items is not large. Thus, it is desirable not to use arithmetic average when the distribution is unevenly spread.

Q32. How to calculate arithmetic mean for individual series ?

Ans :

Individual Series

The process of computing mean in case of individual observations (i.e., where frequencies are not given) is very simple. Add together the various values of the variable and divide the total by the number of items. Symbolically :

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

(OR)

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

Here

X = Arithmetic Mean

ΣX = Sum of all the values of the variable X, i.e., $X_1, X_2, X_3, ...X_n$

N = Number of observations.

Steps

The formula involves two steps in calculating mean :

i) Add together all the values of the variable X and obtain the total, i.e., XX.

ii) Divide this total by the number of the observations, *i.e., N.*

Shortcut method

The arithmetic mean can be calculated by using what is known as an arbitrary origin. When deviations are taken from an arbitrary origin, the formula for calculating arithmetic mean is

$$\overline{X} = A + \frac{\Sigma d}{N} Cx$$

A is the assumed mean and

d is the deviation of items from assumed mean

i.e., d = (X – A*).

Steps

1. Take an assumed mean.
2. Take the deviations of items from the assumed mean and denote these deviations by d.
3. Obtain the sum of these deviations,
4. Apply the formula : $X = A + \frac{\sum d}{N}$.

Q33. How to calculate Arithmetic mean for discrete series ?

Ans :

Calculation of Arithmetic Mean

In discrete series arithmetic mean may be computed by applying

i) Direct method (or)

ii) Short-cut method

i) Direct method

The formula for computing mean is $\overline{X} = \frac{\sum fX}{N}$

Where,

f = Frequency;

X = The variable in question;

N* = Total number of observations, i.e., Σf.

Steps

i) Multiply the frequency of each row with the variable and obtain the total If ΣX.

ii) Divide the total obtained by step (i) by the number of observations, i.e., total frequency.

ii) Short-cut Method

According to this method,

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

where A = Assumed mean; d = (X – A); N - Total number of observations, i.e., Σf.

Steps

i) Take an assumed mean.

ii) Take the deviations of the variable X from the assumed mean and denote the deviations by *d.*

iii) Multiply these deviations with the respective frequency and take the total Σfd.

iv) Divide the total obtained in third step by the total frequency.

Q34. How to calculate arithmetic mean for continuous series ?

Ans :

Calculation of Arithmetic Mean

In continuous series, arithmetic mean may be computed by applying any of the following methods:

i) Direct method,

ii) Short-cut method.

i) Direct Method

When direct method is used

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

where

m = Mid-point of various classes

f = The frequency of each class

N = The total frequency.

Steps

i) Obtain the mid-point of each class and denote it by m.

ii) Multiply these mid-points by the respective frequency of each class and obtain the total Σfm.

iii) Divide the total obtained in step (i) by the sum of the frequency, i.e., N.

ii) Short-cut Method

When short-cut method is used, arithmetic; computed by applying the following formula :

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

where A - assumed mean; d = deviations of mid-points from assumed mean, i.e., (m - A): N = total number of observations.

Steps

i) Take an assumed mean.

ii) From the mid-point of each class deduct the assumed mean.

iii) Multiply the respective frequencies of each class by these deviations and obtain the total Σfd.

iv) Apply the formula: $X = A + \frac{\sum fd}{N}$

PROBLEM

10. Calculate the arithmetic mean of the monthly incomes of the families in a certain locality in Delhi.

Family	A	B	C	D	E	F	G	H	I	J
Income (f)	85	70	10	75	500	8	42	250	40	36

Sol :

$$\text{Arithmetic Mean } (\bar{X}) = \frac{\sum X}{N} = \frac{85+70+10+75+500+8+42+250+40+36}{10}$$

$$= 1116/10 = 111.6$$

∴ The arithmetic mean of the monthly incomes in the locality is Rs. 111.6.

11. Calculate arithmetic mean of the following data by direct, short-cut methods :

Roll No.	1	2	3	4	5	6	7	8	9	10
Marks (Out of 100)	50	65	37	29	92	43	81	36	52	45

Sol :

Calculation of Arithmetic mean by

i) Direct Method

$$\text{Arithmetic mean } (\bar{X}) = \frac{\sum X}{N} = \frac{50+65+37+29+92+43+81+36+52+45}{10} = \frac{530}{10} = 53$$

∴ Average marks of the class = 53 marks.

ii) Short Cut Method

X	X – A = d
50	– 42
65	– 27
37	– 55
29	– 63
92 A	0
43	– 49
81	– 11
36	– 56
52	– 40
45	– 47
	– 390

$$\bar{X} = A + \frac{\sum d}{N} = 92 + \frac{-390}{10}$$

$$= 92 + (-39) = 53.$$

12. Calculate Arithmetic mean from the following data.

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90	90 - 100
No. of Students	33	53	108	221	153	322	439	526	495	50

Sol :

Calculation of Arithmetic mean

Marks	No. of students (f)	Midvalue (M)	fM
0-10	33	5	165
10-20	53	15	795
20-30	108	25	2700
30-40	221	35	7735
40-50	153	45	6885
50-60	322	55	17,710
60-70	439	65	28,535
70-80	526	75	39,450
80-90	495	85	42,075
90-100	50	95	4750
	$\Sigma f = 2400$		$\Sigma fM = 1,50,800$

$$\text{Arithmetic mean} = \frac{\Sigma fM}{\Sigma f} = \frac{1,50,800}{2400} = 62.83$$

3.4.2 Mode

Q35. Define mode.

Ans :

Meaning

Mode may be defined as the value that occurs most frequently in a statistical distribution or it is defined as that exact value in the ungrouped data if each sample which occurs most frequently.

Definitions

(i) According to Croxton and Cowden, "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values."

(ii) According to A.M. Tuttle, "Mode is the value which has the greatest frequency density in its immediate neighbourhood."

(iii) According to Zizek, "The mode is the value occurring most frequently in a series of items and around which the other items are distributed most densely."

Every distribution cannot have a unique value of Mode. It can have two or even more than two modal values. Such distributions are known as Uni-Modal, Bimodal and Multi Modal.

Its graphical representation is given below.

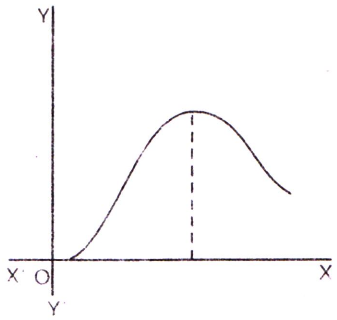


Fig.: Unimodel

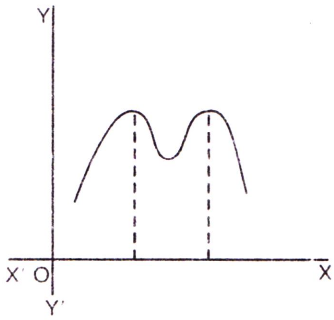


Fig.: Bimodal

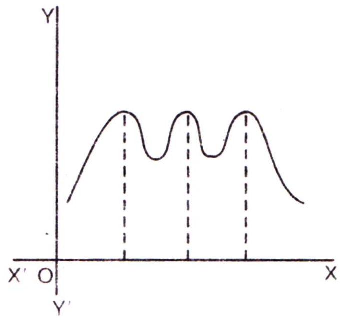


Fig.: Multimodal

Q36. Explain the merits and demerits of Mode.

Ans :

Merits

1. Mode is easy to understand and calculate.
2. It is not influenced much by items on the extremes.
3. It can be located even if the class-intervals are of unequal magnitudes, provided the modal class and the preceding and succeeding it are of the same magnitude.
4. It can be computed for distributions which have open end classes.
5. Mode is not an isolated value like the median. It is the term that occurs most in the series.
6. Mode is not a fictional value that is not found in the series.
7. It can be determined by inspection.
8. It can be located graphically
9. It has wide business application

Demerits

1. Calculation of Mode does not consider all the items of the series. Thus, it is not fully representative of the entire data.
2. It is not rigidly defined.
3. It is not capable of further mathematical treatment.
4. Mode is sometimes indeterminate. There may be 2 (Bi-modal) or more (Multimodal) values.
5. Mode is significantly impacted by fluctuations of sampling. Hence, it is less reliable.
6. Mode is considerably influenced by the choice of grouping. A change in the size of the class interval will change the value of the mode". It is a very unstable average and its true value is difficult to determine.

Q37. How mode is calculated for individual and discrete series.

Ans :

Determination of Mode in Individual Series

The steps involved in calculating mode for individual series are as follows,

1. First arrange the data in ascending or descending order.
2. Check which value is repeating maximum number of times. The value repeating maximum number of times is considered as the value of mode.

Determination of Mode in Discrete Series

In discrete series the value of mode can be determined in two ways,

(i) Inspection

(ii) Grouping and analysis table.

(i) Inspection: Steps

(a) Maximum value or highest value is selected from frequency column

(b) The value corresponding to highest frequency value is considered as mode.

(ii) Grouping and Analysis Table

If there exists an error of maximum frequency, frequency has small value preceding or succeeding it and items are highly focused on any one side then under such case grouping and analysis table are prepared.

Grouping Table

It consists of six columns as follows,

- In column 1-highest frequency is selected and highlighted with a circle or a tick (✓) marks.
- In column 2-frequency values are grouped in two's
- In column 3-Ignore first frequency and group the remaining in two's.
- In column 4-frequencies are grouped in three's
- In column 5-Ignore first frequency and group the remaining frequency values in three's.
- In column 6-Ignore first two frequencies and group remaining frequencies in three's

After preparing the grouping table the maximum value in each column should be marked with a circle.

Analysis Table

- Analysis table is prepared by taking the column number on left side and probable values of mode on right side of the table.
- The values of variable corresponding to the highest frequencies are taken at the top of the analysis table and probable values of mode are determined.
- Finally, the maximum marked variable is considered as 'Mode'.

Q38. How mode is calculated for continuous series.

Ans :

The steps involving in calculating mode in continuous series are as follows,

1. In continuous series the class of model class can be determined in two ways,
   (i) By Inspection (or)
   (ii) By preparing grouping and analysis table.
2. Value of mode is ascertained by using the following formula,

$$\boxed{\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C}$$

Where,

L = Lower limit of model class

f_1 = Frequency of model class

$\Delta_1 = f - f_1$

$\Delta_2 = f - f_2$

C = Class interval.

Problems on Mode

13. Compute the modal value for the following observations and give your reasons.

10, 12, 13, 10, 18, 16, 15, 10, 11, 17, 10, 16, 11, 10, 12, 18, 19, 20, 15, 9

Sol :

Arrange the data in ascending order,

Observations	No.of times repeated (or) appeared
9	1
10	5
11	2
12	2
13	1
15	2
16	2
17	1
18	2
19	1
20	1

Among of all the above number 10 is repeated at maximum i.e. 5 times than the other numbers. Hence mode = 10.

14. Find the mode of the following data.

0, 1, 6, 7, 2, 3, 7, 6, 6, 2, 6, 0, 5, 6, 0

Sol :

Observations	No. of times repeated
0	3
1	1
2	2
3	1
5	1
6	5
7	2

Among of all the above number 6 is repeated at maximum times i.e., 5 times.

Hence mode = 6.

15. Data on monthly income of 70 persons are given below. Calculate value of mode.

Income	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Persons	8	14	19	17	12

Sol :

Mode can be calculated using the given formula,

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

Where

$\Delta_1 = f - f_1$

$\Delta_2 = f - f_2$

From the above distribution, 19 is the highest frequency and hence 20 – 30 is the modal class.

	Income	Persons	
	0 – 10	8	
	10 – 20	14	(f_1)
L	20 – 30	19	(f)
	30 – 40	17	(f_2)
	40 – 50	12	

Here,

$L = 20 \quad f = 19 \quad f_1 = 14 \quad f_2 = 17$

$\Delta_1 = f - f_1$

$19 - 14 = 5$

$\Delta_2 = f - f_2$

$19 - 17 = 2$

$$\text{Mode} = 20 + \frac{5}{5+2} \times 10$$

$$= 20 + \frac{5}{7} \times 10$$

$$= 20 + 7.142 = 27.142$$

$\therefore$ The calculated value of mode is ` 27.142.

---

16. Calculate mode from the following data :

Marks No.of	**0 - 10**	**10 - 20**	**20 - 30**	**30 - 40**	**40 - 50**	**50 - 60**	**60 - 70**
Frequency	**5**	**15**	**20**	**20**	**32**	**14**	**14**

Sol :

	Marks	**Frequency**	
	0 - 10	5	
	10 - 20	15	
	20 - 30	20	
	30 - 40	20	f_1
L	40 - 50	32	f
	50 - 60	14	f_2
	60 - 70	14	

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 \times \Delta_2} \times C$$

$\Delta_1 = f - f_1$

$\Delta_1 = 32 - 20 = 12$

$\Delta_2 = f - f_2$

$\Delta_2 = 32 - 14 = 18$

$$40 + \frac{12}{12+18} \times 10$$

$$40 + \frac{120}{30}$$

$$40 + 4 = 44$$

3.4.3 Median

Q39. Define median. What are the characteristics of median?

Ans :

Meaning

If a group of N observations is arranged in ascending or descending order of magnitude, then the middle value is called median of these observations and is denoted by M.

That is, $M = \frac{N+1}{2}$ th observation.

Definition

According to Croxton and Cowden, "The median is that value which divides a series so that one half or more of the items are equal to or less than it and one half or more of the items are equal to or greater than it."

Characteristics

- Unlike the arithmetic mean, the median can be computed from open-ended distributions. This is because it is located in the median class-interval, which would not be an open-ended class.
- The median can also be determined graphically whereas the arithmetic mean cannot be ascertained in this manner.
- As it is not influenced by the extreme values, it is preferred in case of a distribution having extreme values.
- In case of the qualitative data where the items are not counted or measured but are scored or ranked, it is the most appropriate measure of location.

Q40. What are the advantages and disadvantages of median?

Ans :

Advantages

(i) The median, unlike the mean, is unaffected by the extreme values of the variable.

(ii) It is easy to calculate and simple to understand, particularly in a series of individual observations and a discrete series. .

(iii) It is capable of further algebraic treatment. It is used in calculating mean deviation.

(iv) It can be located by inspection, after arranging the data in order of magnitude.

(v) It can be determined graphically.

(vi) Median can be calculated in case of open-end classes.

(vii) Median is defined rigidly.

Disadvantages

(i) For calculation, it is necessary to arrange the data, other averages do not need any such arrangement.

(ii) It is amenable to algebraic treatment in a limited sense. Median cannot be used to calculate the combined median of two or more groups, like mean.

(iii) Median is affected more by sampling fluctuations than the mean.

(iv) It is not based on all observations. So it is a positional average.

Q41. How median is calculated for individual series ?

Ans :

(i) Arrange the data in ascending or descending order of magnitude. (Both arrangements would give the same answer.)

(ii) In a group composed of an odd number of values such as 7, add 1 to the total number of values and divide by 2. Thus, $7+1$ would be 8 which divided by 2 gives 4 the number of the value starting at either end of the numerically arranged groups will be the median value.

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

(iii) In a group composed an even number of values, then median $= \frac{N}{2}+1^{th}$ item

(or)

$$\frac{\text{Two middle values}}{2}$$

Q42. How median is calculated for discrete series ?

Ans :

(i) Arrange the data in ascending or descending order of magnitude,

(ii) Find out the cumulative frequencies.

(iii) Apply the formula: Median $=$ Size of $\left(\frac{N+1}{2}\right)^{th}$ item

(iv) Now look at the cumulative frequency column and find that total which is either equal to Size of $\frac{N+1}{2}$ or next higher to that and determine the value of the variable corresponding to it. That gives the value of median.

Q43. How median is calculated for continuous series ?

Ans :

1. Arrange the given data in ascending order
2. Calculate cumulative frequency
3. Find N_1 as N_1 = N/2 for Median
4. In C.f find the value equal to or just greater than N_1
5. Find f and L just one step below the C.f

$$\text{Median} = L + \frac{\frac{N}{2}-F}{f} \times C$$

L = Lower limit of the median class,

F = Cumulative frequency of the class preceding the median class (or) sum of the frequencies of all classes lower than the median class.

f = Simple frequency of the median class

C = The class interval of the median class.

PROBLEMS ON MEDIAN

17. Calculate Median from the following.

61	**62**	**63**	**61**	**63**	**64**	**64**	**60**	**65**	**63**	**64**	**65**	**66**	**64**

Sol :

Step 1

Arrange the data in ascending order for the given i.e.,

60, 61, 61, 62, 63, 63, <u>63, 64,</u> 64, 64, 65, 65, 65, 66

Step 2

Apply the formulae for :

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

$$= \frac{14+1}{2} = \text{Size of 7.5th item}$$

when 7.5 is equalled to 7th and 8th items of the data. Hence 7^{th} = 63, 8th item = 64

$$\text{median} = \frac{63+64}{2} = 63.5$$

18. Calculate median from the following.

22, 26, 14, 30, 18, 11, 35, 41, 12, 32

Sol :

Step I : Arrange the data in ascending order i.e.,

11, 12, 14, 18, <u>22, 26,</u> 30, 32, 35, 41

Step II : Calculation of given ascending to formulae

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

$$= \frac{10+1}{2} = \frac{11}{2} = \text{Size of 5.5th item}$$

$$\text{Median} = \frac{\text{5th item} + \text{6th item}}{2} = \frac{22+26}{2} = \frac{48}{2} = 24.$$

19. From the following data, find the value of median.

Income (`)	200	250	130	270	300	230
No. of Persons	34	36	26	30	16	40

Sol :

Step I : Arrange the data in ascending order.

Income (Rs.) (x)	No. of Persons (F)	Cumulative Frequency (CF)
130	26	26
200	34	60
230	40	100
250	36	136
270	30	166
300	16	182
	N = 182	

Step II: Apply the formulae for determination of media.

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

$$\text{Median} = \text{Size of } \frac{182+1}{2}\text{th item}$$

$$\text{Median} = \text{Size of } \frac{183}{2} = 91.5 \text{ item}$$

Median Size of 91.5 is representing in 100 at cumulative frequency which is representing in corresponding (x) column is 230 in income level.

$\therefore$ Median = 230.

20. From the following data, compute Median

Class Interval	0-10	10-20	20-30	30-40	40-50
Frequency	15	100	170	120	40

Sol :

Marks	Frequency	Cumulative Frequency (CF)
0 - 10	15	15
10 - 20	100	115 CF
L 20 - 30	170 f	285
30 - 40	120	405
40 - 50	40	445

$$\text{Median} = L + \frac{\frac{N}{2} - F}{f} \times C$$

$$\text{Median class interval} = \text{Size of } \frac{N}{2}\text{th item}$$

$$\text{Median class interval} = \text{Size of } = \frac{445}{2} = 222.50 \text{ item}$$

∴ 222.50th item consisting in CF 285 which is representing in corresponding class interval is 20-30.

$$\text{Median} = 20 + \left[\frac{\frac{N}{2} - F}{f}\right] \times C = 20 + \left(\frac{222.50 - 115}{170}\right) \times 10$$

$$\text{Median} = 20 + 0.632 \times 10$$

$$\text{Median} = 20 + 6.32$$

$$\text{Median} = 26.32$$

3.4.4 Relationship among Mean, Median and Mode

Q44. Describe the relationship between Mean, Median and Mode.

Ans :

Mode touches the peak of the curve indicating maximum frequency. Median divides -the area of the curve in two equal halves and Mean is the centre of gravity. The three values are inter-related. The following points list the relationship between the various averages.

- In a distribution, the relative positions of mean, median, and the mode depend upon the Skewness of the distribution. If the distribution is symmetrical, then Mean = Median = Mode.
- If the distribution is positively skewed (the longer tail of the distribution is towards the right), the Mode will be lesser, than the Median, which in turn will be lower than the Arithmetic Mean. In case of a negatively skewed distribution, the Mode will be greater than the Median, which in turn will be greater than the Arithmetic Mean.
- In a given distribution, if all the observations are positive, Arithmetic Mean is greater than Geometric Mean, which in turn is greater than Harmonic Mean. With a distribution of moderate Skewness, median tends to be approximately 1/3rd as far away from the mean as from the mode.

  Mode = 3 Median-2 Mean
- Mean-Mode = 3 (Mean-Median). Thus, the difference between Mean and Mode is three times, the difference between mean and median.

PROBLEMS

21. If the Mode and Mean of a moderately skewed series are 30.2 and 20.4 respectively, what would be its Median?

Sol :

We know that 3Median – 2Mean = Mode

Given mode = 30.2 & Mean = 20.4

Median = ?

3Median = Mode + 2 Mean

3 Median = 30.2 + 2(20.4)

3 Median = 30.2 + 40.8

3 Median = 71.0

$$\text{Median} = \frac{71}{3} = 23.6$$

22. If mode = 16, mean = 15.6 find the median.

Sol :

Given that,

Mean = 15.6 and Mode = 16

Mode = 3 Median – 2 Mean

16 – 3 Median – (2 × 15.6)

16 + 31.2 = 3 Median

47.2 = 3 Median

$$\text{Median} = \frac{47.2}{3}$$

Median = 15.733

3.5 DISPERSION

Q45. Define Dispersion. What are the objectives of Measuring Dispersion.

Ans :

Meaning

The concept of dispersion is related to the extent of scatter or variability in observations. The variability, in an observation, is often measured as its deviation from a central value. A suitable average of all such deviations is called the measure of dispersion. Since most of the measures of dispersion are based on the average of deviations of observations from an average, they are also known as the averages of second order.

Definitions

As opposed to this, the measures of central tendency are known as the averages of first order. Some important definitions of dispersion are given below:

(i) **According to A.L. Bowley,** "Dispersion is the measure of variation of the items."

(ii) **According to Connor,** "Dispersion is the measure of extent to which individual items vary."

(iii) **According to Simpson and Kafka,** "The measure of the scatteredness of the mass of figures in a series about an average is called the measure of variation or dispersion."

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

Objectives

The main objectives of measuring dispersion of a distribution are:

1. **To test reliability of an average:** A measure of dispersion can be used to test the reliability of an average. A low value of dispersion implies that there is greater degree of homogeneity among various items and, consequently, their average can be taken as more reliable or representative of the distribution.

2. **To compare the extent of variability in two or more distributions:** The extent of variability in two or more distributions can be compared by computing their respective dispersions. A distribution having lower value of dispersion is said to be more uniform or consistent.

3. **To facilitate the computations of other statistical measures:** Measures of dispersions are used in computations of various important statistical measures like correlation, regression, test statistics, confidence intervals, control limits, etc.

4. **To serve as the basis for control of variations:** The main objective of computing a measure of dispersion is to know whether the given observations are uniform or not.

Q46. Explain the Significance of Measuring Dispersion.

(OR)

Explain the purpose of Dispersion.

Ans :

Measures of variation are needed for four basic purposes :

1. To determine the reliability of an average.
2. To serve as a basis for the control of the variability.
3. To compare two or more series with regard to their variability.
4. To facilitate the use of other statistical measures.

Q47. What are the Characteristics of a good Measures of Dispersion.

Ans :

(i) It should be easy to calculate.

(ii) It should be easy to understand.

(iii) It should be rigidly defined.

(iv) It should be based on all the observations.

(v) It should be capable of further mathematical treatment.

(vi) It should not be unduly affected by extreme observations.

(vii) It should not be much affected by the fluctuations of sampling.

Q48. Explain the types of Measures of Dispersion.

Ans :

There are different measures of dispersion. These measures can be classified into:

i) Absolute measures and

ii) Relative measures.

i) An 'Absolute' measure is one that is expressed in terms of the same unit in which the variable (or given data) is measured.

ii) A 'Relative" measure of dispersion is expressed as a pure number (without any units) which enables comparison of the levels of dispersion from a central tendency across different series (stated in different units). These measures are also called as "Co-efficient(s) of Dispersion". The important methods of studying variation are listed as under:

Absolute measures

(i) The Range

(ii) Inter Quartile Range and Quartile Deviation.

(iii) The Mean Deviation (or) Average Deviation.

(iv) The Standard Deviation and Variance

(v) The Lorenz Curve

Relative measures

(i) Coefficient of Range

(ii) Coefficient of Quartile Deviation.

(iii) Coefficient of Mean Deviation.

(iv) Coefficient of Variation

3.5.1 Range

Q49. What is Range ? Explain.

Ans :

Range is the simplest method of studying dispersion. It is defined as the difference between the value of the smallest item and the value of the largest item included in the distribution. Symbolically,

$$\boxed{\text{Range} = L - S}$$

where

L = Largest item, and

S = Smallest item.

The relative measure corresponding to range, called the coefficient of range, is obtained by applying the following formula :

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

If the averages of the two distributions are about the same, a comparison of the range indicates that the distribution with the smaller range has less dispersion, and the average of that distribution is more typical of the group.

Q50. What are the merits and limitations of Range ?

Ans :

The merits and limitations of Range can be enumerated here.

Merits (or) Advantages

(i) Amongst all the methods of studying dispersion range is the simplest to understand and the easiest to compute.

- It is one of those measures which are rigidity defined.

(ii) It takes minimum time to calculate the value of range. Hence, if one is interested in getting a quick rather than a very accurate picture of variability one may compute range.

- It gives us the total picture of the problem even with a single glance.
- It is used to check the quality of a product for quality control. Range plays an important role in preparing R-charts, thus quality is maintained.

Demerits (or)Limitations (or) Drawbacks

1. Range is not based on all the terms. Only extreme items reflect its size. Hence range cannot be completely representative of the data as all other middle values are ignored.
2. Due to above reason range is not a reliable measure of dispersion.
3. Range does not change even the least even if all the terms and variables are changed.
4. Range is too much affected by fluctuation of sampling. Range changes from sample to sample. As the size of sample increases range increases and vice versa.
5. It does not tell us anything about the variability of other data.
6. For open-end intervals, range is indeterminate because lower and upper limits of first and last interval are not given.

Q51. What are the uses of range ?

Ans :

(i) Quality Control

The object of quality control is to keep a check on the quality of the product without 100% inspection. When statistical methods of quality control are used, control charts are prepared and in preparing these charts range plays a very important role. The idea basically is that if the range - the difference between the largest and smallest mass produced items-increases beyond a certain point, the production machinery should be examined to find out why the items produced have not followed their usual more consistent pattern

(ii) Fluctuations in the Share Prices

Range is useful in studying the variations in the prices of stocks and shares and other commodities that are sensitive to price changes from one period to another.

(iii) Weather Forecasts

The meteorological department does make use of the range in determining, say, the difference between the minimum temperature and the maximum temperature. This information is of great concern to the general public because they know as to within what limits the temperature is likely to vary on a particular day.

(iv) Everyday Life

The range is a most commonly used measure of dispersion in everyday life. Questions of the form "what is the minimum and maximum temperature on a particular day"? "What is the difference between the wages earned by workers of a particular factory" ? How much one spends on petrol in his car/scooter in a month"- are all usually answered in the form of range. Answers to questions such as these are usually given in the form of 'Between such and such]. Regardless of the crudity of expression the answer is still a range.

PROBLEMS ON RANGE

23. The following are the wages of 10 workers of a factory. Find the range of variation and also compute the coefficient of range.

275, 200, 370, 240, 100, 290, 400, 500, 180, 350

Sol :

Step I

Arrange the data in ascending order i.e.

100, 180, 200, 240, 275, 290, 350, 370, 400, 500

Range = L – S

where as = L = 500, S = 100

Range (R) = 500 – 100 = 400

Coefficient of range = $\frac{L-S}{L+S} = \frac{500-100}{500+100}$

Coefficient of range = $\frac{400}{600}$ = 0.667.

24. The following are the marks of 100 students of a class. Find the range of variation of marks and also compute coefficient of range.

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students	8	16	26	20	12	10	8

Sol :

From the given data

Highest (L) marks is 80 and

Lowest (S) marks is 10. Hence

Range (R) = L – S

(R) = 80 – 10 = 70

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

Coefficient of Range = $\frac{80-10}{80+10}$

Coefficient of Range = $\frac{70}{90}$

Coefficient of Range = 0.77.

3.5.2 Quartile Deviation

Q52. What is Quartile Deviations ?

Ans :

It is based on two extreme items and it fails to take account of the scatter within the range. From this there is reason to believe that if the dispersion of the extreme items is discarded, the limited range thus established might be more instructive.

For this purpose there has been developed a measure called the interquartile range, the range which includes the middle 50 per cent of the distribution. That is, one quarter of the observations at the lower end, another quarter of the observations at the upper end of the distribution are excluded in computing the interquartile range. In other words, interquartile range represents the difference between the third quartile and the first quartile.

$$\boxed{\text{Inter quartile range } = Q_3 - Q_1}$$

Very often the interquartile range is reduced to the form of the Semi-interquartile range or quartile deviation by dividing it by 2.

$$\boxed{\text{Quartile Deviation or Q.D. } = \frac{Q_3 - Q_1}{2}}$$

Coefficients of Quartile Deviation

Quartile deviation is an absolute measure of dispersion. Its relative measure is the coefficient of Quartile deviation.

Coefficients of Quartile Deviation

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

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

It is used to compare the degree of variation in different series.

Q53. Explain merits and demerits of quartile deviations.

Ans :

Merits

1. It is the simple to calculate and very easy to understand.
2. It is not impacted by extreme values
3. It can be computed for open ended distributions and for data containing unequal classes

Demerits

1. It is impacted by sample size and composition of the sample.
2. It is not amenable to algebraic or statistical treatment
3. It does not tell us anything about the spread of the various data items across the measure of central tendency.

Q54. How to calculate quartile deviations?

Ans :

Low Quartile (Q_1),

$$= \text{Size of } \frac{N+1}{4} \text{th item}$$

N = No. of observations

Upper Quartile (Q_3),

$$= \text{Size of } \frac{3(N+1)}{4} \text{th item.}$$

Discrete Series

i) $Q_1 = \text{Size of } \frac{N+1}{4}\text{th item}$

ii) $Q_3 = \text{Size of } \frac{3(N+1)}{4}\text{th item}$

iii) Quartile Deviation (Q.D) = $\frac{Q_3 - Q_1}{2}$

Continuous Series

$$Q_1 = L + \frac{\frac{N}{4} - F}{f} \times C$$

$$Q_3 = L + \frac{\frac{3N}{4} - F}{f} \times C$$

PROBLEMS ON QUARTILE DEVIATION

25. Find out the value of quartile deviation and its coefficient for the following data :

Roll No	1	2	3	4	5	6	7
Marks	25	33	45	17	35	20	55

Sol :

Calculation of Quartile Deviation

Marks arranged in ascending order 17 20 25 33 35 45 55

$$Q_1 = \text{Size of } \left[\frac{N+1}{4}\right]^{th} \text{ Item}$$

N = No. of observations.

$$= \text{Size of } \left[\frac{7+1}{4}\right]^{th} \text{ Item} = 2^{nd} \text{ Item} = 20$$

$$Q_3 = \text{Size of } 3\left[\frac{N+1}{4}\right]^{th} \text{ Item}$$

$$= \text{Size of } 3\left[\frac{7+1}{4}\right]^{th} \text{ Item} = 6^{th} \text{ Item} = 45$$

$$\text{Quartile Deviation (Q.D)} = \frac{Q_3 - Q_1}{2} = \frac{45-20}{2} = \frac{25}{2} = 12.5$$

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{45-20}{45+20} = \frac{25}{65} = 0.38.$$

26. Find out the value of quartile deviation from the following data

Roll No.	1	2	3	4	5	6	7
Marks :	30	42	60	18	45	24	75

Sol :

Rearrange into ascending order

x = 18, 24, 30, 42, 45, 60, 75

$$\text{Quartile deviation } Q_1 = \text{size of } \left(\frac{N+1}{4}\right)^{th} \text{ item}$$

$$\text{Size of } \left(\frac{7+1}{4}\right)^{th} \text{ item} = 2^{nd} \text{ item}$$

The size of 2^{nd} item is 24 i.e. $Q_1 = 24$

$$Q_3 = \text{size of } 3\left(\frac{7+1}{4}\right)^{th} \text{ item}$$

$$= 6^{th} \text{ item} = 60, \text{ then } Q_3 = 60$$

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

$$= \frac{60-24}{2} = \frac{36}{2} = 18$$

$$Q.D = 18$$

27. Calculate the Quartile Deviation and Coefficient of Quartile Deviation of the following distribution.

Marks (X)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No.of Students (f)	11	18	25	28	30	33	22	15	12	10

Sol :

Calculation of Quartile Deviation

X	f	C.f
0-10	11	11
10-20	18	29 F
L 20-30	25 f	54 (Q_1)
30-40	28	82
40-50	30	112
50-60	33	145 F
L 60-70	22 f	167 (Q_3)
70-80	15	182
80-90	12	194
90-100	10	204

Calculation of Q_1

$$Q_1 = \left(\frac{N}{4}\right)^{th} \text{ Item}$$

$$= \left(\frac{204}{4}\right)^{th} \text{ Item}$$

$$= 51^{th} \text{ Item}$$

Size of 51^{th} Item lies in CF 54. So Q_1 class interval = 20–30

$$Q_1 = L + \frac{\frac{N}{4} - F}{f} \times C$$

$$= 20 + \frac{51 - 29}{25} \times 10$$

$$= 20 + \frac{22 \times 10}{25}$$

$$= 20 + \frac{220}{25}$$

$$= 20 + 8.8 = 28.8.$$

Calculation of Q_3

$$= Q_3 = 3\left(\frac{N}{4}\right)^{th}$$

$$= 3\left(\frac{204}{4}\right)^{th} \text{ Item } = 3(51)^{th} \text{ Item } = 153^{th} \text{ Item}$$

Size of 153^{th} Item lies in CF 167. So Q_3 class interval = 60 – 70

$$Q_3 = L_3 + \frac{\frac{3N}{4} - F}{f} \times C$$

$$= 60 + \frac{153 - 145}{22} \times 10$$

$$= 60 + \frac{80}{22}$$

$$= 60 + 3.63$$

$$= 63.63$$

Calculation of Quartile Deviation

$$\text{Quartile Deviation (Q.D)} = \frac{Q_3 - Q_1}{2} = \frac{63.64 - 28.8}{2} = 17.4$$

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{63.64 - 28.8}{63.64 + 28.8} = \frac{34.84}{92.44} = 0.37$$

3.5.3 Mean Deviation

Q55. Define mean deviation. State the merits and demerits of mean deviation.

Ans :

Meaning

"Mean Deviation of a series is the arithmetic average of the deviations of various items from a measure of central tendency (either mean, median or mode)". Theoretically, deviations can be taken from any of the three averages mentioned above, but in actual practice it is calculated either from mean or from Median. While Calculating deviations algebraic signs are not taken into account.

Merits

(i) It is rigidly defined

(ii) It is not least impacted by sampling fluctuations

(iii) It takes into account every single value in the series

(iv) Mean deviation from Median is least impacted due to extreme values

(v) It is extensively used in multiple fields such as Economics, Commerce, etc as it is the best measure for comparison of two or more series.

Demerits

(i) It is relatively difficult to compute

(ii) It is not amenable to further algebraic or statistical treatment.

(iii) It is difficult for a layman to understand as to why or when a particular average should be considered for calculation of Mean Deviation. The Mean Deviations obtained by taking the Mean, Median and Mode as average differ widely

(iv) It is not effective for open ended series, particularly when the average is Arithmetic Mean.'

Q56. How to calculate mean deviation for individual series ?

Ans :

Individual Series

If X_1, X_2, X_3, X_N are N given observations then the deviation about an average. A is given by

$$\frac{\Sigma|D|}{N}$$

Steps

i) Compute the median of the series.

ii) The deviations of items from median ignoring ± signs and denote these deviations by $|D|$.

iii) Obtain the total of these deviations, i.e., $\Sigma|D|$.

iv) Divide the total obtained in step

v) By the number of observations.

The relative measure corresponding to the mean deviation, called the coefficient of mean deviation, is obtained by dividing mean deviation by the particular average used in computing mean deviation. Thus, if mean deviation has been computed from median, the coefficient of mean deviation shall be obtained by dividing mean deviation by median.

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Median}}$$

If mean has been used while calculating the value of mean deviation, in such a case coefficient of mean deviation shall be obtained by dividing mean deviation by the mean.

Q57. How to calculate mean deviation for discrete series ?

Ans :

In discrete series the formula for calculating mean deviation is

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

$|D|$ denotes deviation from median ignoring signs.

N = Sum of frequency

Steps

i) Calculate the median of the series.

ii) Take the deviations of the items from median ignoring signs and denote them by |D|.

iii) Multiply these deviations by the respective frequencies and obtain the total $\Sigma f|D|$

iv) Divide the total obtained in Step (ii) by the number of observations. This gives us the value of mean deviation.

Q58. How to calculate mean deviation for continuous series ?

Ans :

For calculating mean deviation in continuous series the procedure remains the same as discussed above. The only difference is that here we have to obtain the mid-point of the various classes and take deviations of these points from median. The formula is same, i.e..

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

Steps

i) Calculate the median of the series.

ii) Take the deviations of the items from median ignoring signs and denote them by |D|.

iii) Multiply these deviations by the respective frequencies and obtain the total Sf|D|.

iv) Divide the total obtained in step (ii) by the number of observations. This gives us the value of mean deviations.

PROBLEMS

28. Calculate mean deviation of the following from mean and median

2, 6, 11, 14, 16, 19, 23

Sol. :

X	$X - \overline{X} = \|D\|$	X – Median = |D|
2	11	12
6	7	8
11	2	3
14	1	0
16	3	2
19	6	5
23	10	9
91	**40**	**39**

(a) Mean $(\overline{X}) = \frac{\Sigma x}{N} = \frac{91}{7} = 13$

MD through mean $= \frac{\Sigma|D|}{N} = \frac{40}{7} = 5.71$

(b) Median $= \text{size of} \left(\frac{N+1}{N}\right)^{th}$ item

$= \text{size of} \left(\frac{7+1}{2}\right)$

$= 8/2 = 4^{th}$ observation $= 14$

(c) Median = size of 4th Item = 14

Mean Deviation (MD) through

Median $= \frac{\Sigma|D|}{M}$

Mean Deviation (MD) $= \frac{39}{7}$

Mean Deviation (MD) = 5.57

29. Calculate the mean deviation and its coefficient from the following data.

Income of Members (`)
1,000
1,500
2,500
3,000
3,500
4,000

Sol :

Calculation of Mean Deviation

Income of Members (x)	**Deviation from Median |D| = x – A (2500)**
1,000	1,500
1,500	1,000
2,000	500
2,500	0
3,000	500
3,500	1,000
4,000	1,500
N = 7	**Σ|D| = 6,000**

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

$$= \frac{7+1}{2} = \frac{8}{2} = 4^{th} \text{ item,}$$

$$\text{Median deviation} = \frac{\Sigma|D|}{N}$$

Where,

$|D|$ = Deviation from median ignoring signs.

$$\text{M.D} = \frac{6,000}{7} = 857.14$$

$$\therefore \quad \text{Coefficient of M.D} = \frac{857.14}{2500} = 0.343$$

30. From the following data calculate mean deviation from the median.

C.I.	16 - 20	21 - 25	26 - 30	31 - 35	36 - 40	41 - 45	46 - 50	51 - 55	56 - 60
Frequency	8	15	13	20	11	7	3	2	1

Sol :

Calculation of Mean Deviation from Median

C.I	Frequency (F)	C.F	Mid-Value (m)	$\|M - A\| = \|D\|$ $\|M - 31.5\|$	$F\|M - A\| = f\|D\|$
16 - 20	8	8	18	13.5	108.0
21 - 25	15	23	23	8.5	127.5
26 - 30	13	36 F	28	3.5	45.5
L 31 - 35	20 f	56	33	1.5	30.0
36 - 40	11	67	38	6.5	71.5
41 - 45	7	74	43	11.5	80.5
46 - 50	3	77	48	16.5	49.5
51 - 55	2	79	53	21.5	43.0
56 - 60	1	80	58	26.5	26.5
	N = 80				**$\Sigma f\|D\|$ = 582**

$$\text{Median} = \text{size of } \frac{N}{2}^{th} \text{ item}$$

$$= \frac{80}{2} = 40^{th} \text{ item}$$

C.F is just greater than 40 is 56

Median lies in the class 31-35. However, the real limit of this class under exclusive method is 30.5 – 35.5.

$$\text{Median} = L + \frac{\frac{N}{2} - F}{f} \times C$$

$$L = 30.5, F = 36, f = 20 \quad C = 5$$

$$= 30.5 + \frac{40-36}{20} \times 5$$

$$= 30.5 + \frac{4}{20} \times 5$$

$$= 30.5 + [0.2] \times 5$$

$$= 30.5 + 1 = 31.5$$

$$\text{M.D from Median} = \frac{\Sigma f|D|}{N} = \frac{582}{80} = 7.275$$

3.5.4 Standard Deviation

Q59. What is standard deviation ? Explain its merits and demerits.

Ans : **(Imp.)**

Meaning

The standard deviation concept was introduced by Karl Pearson in 1823. It is by far the most important and widely used measure of studying dispersion. Its significance lies in the fact that it is free from those defects from which the earlier methods suffer and satisfies most of the properties of a good measure of dispersion. Standard deviation is also known as root mean square deviation for the reason that it is the square root of the mean of the squared deviation from the arithmetic mean. Standard deviation is denoted by the small Greek letter σ (read as sigma).

The standard deviation measures the absolute dispersion (or variability of distribution; the greater the amount of dispersion or variability), the greater the standard deviation, the greater will be the magnitude of the deviations of the values from their mean. A small standard deviation means a high degree of uniformity of the observation as well as homogeneity of a series; a large standard deviation means just the opposite.

Merits

1. It is rigidly defined
2. It takes into account every single value in the series
3. It is amenable to further algebraic or statistical treatment.
4. It is extensively used in various other statistical calculations such as correlation, regression, sampling etc

Demerits

1. It is relatively difficult to compute
2. It is calculated with only Arithmetic Mean as the average. Standard deviation from other averages such as Median is not an effective measure of dispersion.

Q60. How to compute the standard deviations for individual series ?

Ans :

Individual Series

In case of individual observations standard deviations may be calculate by applying any of the following two methods, i.e.

i) By taking deviations of the items from the actual mean.

ii) By taking deviations of the items from an assumed mean.

Deviation taken from actual mean : when deviations taken from actual mean the following formula is applied.

$$\sigma = \sqrt{\frac{\sum x^2}{N}}$$

where as

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

N = Total number of observations.

Steps

i) Calculate the actual for the series i.e. $\bar{X}$

ii) Find X deviation from $\bar{X}$ i.e. $x = (X - \bar{X})$

iii) Square these deviations and obtain the total Σx^2

iv) Divide ΣX^2 by the total number of observations and extract the square root.

Q61. How to compute the standard deviations for discrete series ?

Ans :

For calculating standard deviation in discrete series, any of the following methods may be applied:

(a) Actual mean method.

(b) Assumed mean method.

(c) Step deviation method.

(a) Actual Mean Method. When this method is applied, deviations are taken from the actual mean, i.e., we find $(X - \bar{X})$ and denote these deviations by x. These deviations are then squared and multiplied by the respective frequencies. The following formula is applied :

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}}, \text{ where } x = (X - \bar{X})$$

N = Sum of observation/Frequency.

However, in practice this method is rarely used because if the actual mean is in fractions the calculations take a lot of time.

(b) Assumed Mean Method. When this method is used, the following formula is applied :

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

where $d = (X - A)$

$N =$ Sum of frequency

Steps

i) Take the deviations of the items from an assumed mean and denote these deviations by d.

ii) Multiply these deviations by the respective frequencies and obtain the total, Sfd.

iii) Obtain the squares of the deviations, i.e., calculate d^2.

iv) Multiply the squared deviations by the respective frequencies, and obtain the total, Sfd^2.

v) Substitute the values in the above formula.

(c) Step Deviation Method: When this method is used we take deviations of midpoints from an assumed mean and divide these deviations by the width of class interval, i.e., 'C'. In case class intervals are unequal, we divide the deviations of midpoints by the lowest common factor and use 'C' in the formula for calculating standard deviation.

The formula for calculating standard deviation is :

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

where,

$$d = \frac{(X - A)}{C} \text{ and } C = \text{class interval.}$$

The use of the above formula simplifies calculations.

Q62. How to compute the standard deviations for continuous series ?

Ans :

In continuous series any of the methods discussed above for discrete frequency distribution can be used. However, in practice it is the step deviation method that is most used. The formula is

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

$$\text{where } d = \frac{(m - A)}{C},\ C = \text{class interval}$$

Steps

i) Find the mid-points of various classes.

ii) Take the deviations of these mid-points from an assumed mean and denote these deviations by d.

iii) Wherever possible take a common factor and denote this column by d.

iv) Multiply the frequencies of each class with these deviations and obtain Σfd.

v) Square the deviations and multiply them with the respective frequencies of each class and obtain Σfd^2.

3.5.5 Coefficient of Variation

Q63. Define Coefficient of Variation.

Ans :

Coefficient of Variation (CV) was proposed by Karl Pearson. It is used to compare the variability of two (or) more distributions. A distribution with greater C.V. is considered as more variable or less consistent, less unit form, less stable or less homogeneous distribution and the distribution with less C.V is considered as less variable or more consistent, more uniform, more stable or more homogeneous distribution.

$$\text{Coefficient of variation} = \frac{\text{Standard deviation}}{\text{Arithmetic mean}} \times 100$$

PROBLEMS ON STANDARD DEVIATION

31. Calculate standard deviation for the following data:

Sl. No.	Weekend Income (`)
1	270
2	350
3	258
4	282
5	218
6	202
7	364
8	184

Sol :

Calculation of Standard Deviation

Sl.No.	X	$x = [X - \bar{X}]$	x^2
1	270	4	16
2	350	84	7056
3	258	– 8	64
4	282	16	256
5	218	– 48	2304
6	202	– 64	4096
7	364	98	9604
8	184	– 82	6724
	$\Sigma X = 2128$		$\Sigma X^2 = 30120$

(i) Arithmetic Mean ($\bar{X}$) $= \dfrac{\Sigma X}{N} = \dfrac{2128}{8}$

$= 266$

(ii) Standard Deviation (σ) $= \sqrt{\dfrac{\Sigma x^2}{N}}$

$= \sqrt{\dfrac{30120}{8}}$

$= \sqrt{3765}$

$= 61.35$

32. From the following information calculate standard deviation.

X	4.5	14.5	24.5	34.5	44.5	54.5	64.5
F	1	5	12	22	17	9	4

Sol :

Calculation of Standard Deviation

X	f	X – A = d	d²	fd	fd²
4.5	1	–30	900	–30	900
14.5	5	–20	400	–100	2000
24.5	12	–10	100	–120	1200
34.5 A	22	0	0	0	0
44.5	17	10	100	170	1700
54.5	9	20	400	180	3600
64.5	4	30	900	120	3600
	70			220	13,000

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

$\sigma = \sqrt{\dfrac{13000}{70} - \left(\dfrac{220}{70}\right)^2}$

$\sigma = \sqrt{185.71 - (3.14)^2}$

$\sigma = \sqrt{185.71 - 9.8596}$

$\sigma = \sqrt{175.85}$

$\sigma = 13.26$

33. Calculate Standard Deviation and Coefficient of Variation from the following data:

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
No. of Students	5	7	14	28	12	9	6	2

Sol :

Calculation of Standard Deviation

Marks	Frequency	Midvalue	d	d^2	fd^2	fd
0 - 10	5	5	– 3	9	45	– 15
10 - 20	7	15	– 2	4	28	– 14
20 - 30	14	25	– 1	1	14	– 14
30 - 40	28	35 A	0	0	0	0
40 - 50	12	45	1	1	12	12
50 - 60	9	55	2	4	36	18
60 - 70	6	65	3	9	54	18
70 - 80	2	75	4	16	32	8
	83				221	13

$$\text{Mean} = A + \frac{\Sigma fd}{N} \times C$$

$$= 35 + \frac{13}{83} \times 10$$

$$= 35 + 15 = 36.5$$

$$\text{S.D} = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C$$

$$= \sqrt{\frac{221}{83} - \left(\frac{13}{83}\right)^2} \times 10$$

$$= \sqrt{2.66 - 0.024} \times 10$$

$$= \sqrt{2.636} \times 10$$

$$= 1.623 \times 10$$

$$= 16.23$$

$$\text{Coefficient of variation} = \frac{\text{S.D}}{\text{Mean}} \times 100$$

$$= \frac{16.23}{36.5} \times 100$$

$$= 0.44 \times 100 = 44.46$$

3.6 Small Sample Tests, t-Distribution

Q64. Define Small sample tests and t-distribution.

(OR)

When is t-test used?

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

Meaning

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.

Distribution

When population standard deviation (σ_y) is not known and the sample is of small size (i.e., $n \leq 30$), we use 't' distribution (student's 't' distribution) for the sampling distribution of mean and workout 't' variable as,

$$t = \frac{(\overline{x} - \mu)}{\frac{\sigma_2}{\sqrt{n}}}$$

Where, $$\sigma_s = \sqrt{\frac{\Sigma(x_i - \overline{x})^2}{n-1}}$$

i.e., the sampling standard deviation

r-distribution is also symmetrical and is very close to the distribution of standard normal variate, Z, except for small values of 'n' There are different Y distributions one for each sample size i.e., for different degrees of freedom. The degrees of freedom for a sample of size *n* is (n – 1).

3.6.1 Properties and Applications

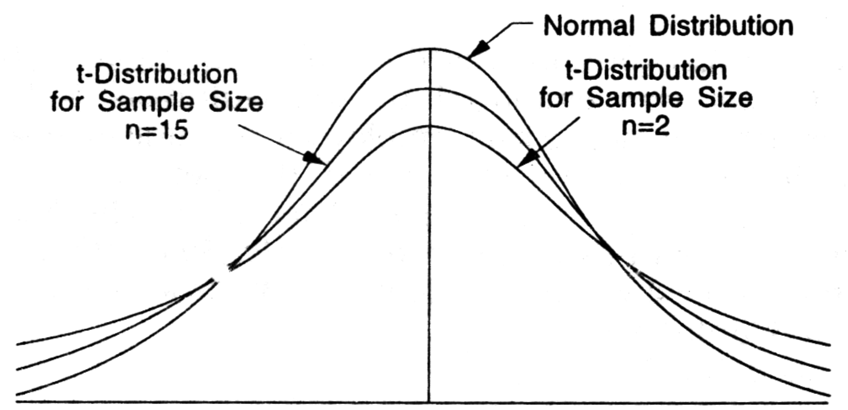
Q65. Explain the properties and applications of t-distribution.

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

Properties

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 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.

3.6.2 Testing for One and Two Means

Q66. Explain the test concerning the significance of single sample mean.

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

To test if sample mean $\overline{x}$ differs significantly from the hypothetical value m of the population mean, we use the following t-test.

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

Where,

n = Sample Size

$n = \frac{1}{n} \sum x_1$ (Sample mean)

μ = Population mean from which sample is taken

$s = \sqrt{\frac{1}{n-1}\Sigma(x_1 - \overline{x})^2}$ (S.D. Sample)

Degree of freedom = n – 1

Decision

If calculate value of $|t| \geq$ table value for (n–1) d.f. at 5% level then reject the hypothesis and say that the two means differ significantly.

The sample does not seem to have been drawn from the population with mean μ.

If calculated value of $|t| <$ table value for (n – 1) d.f. at 5% level then accept the hypothesis.

PROBLEMS

34. A soap manufacturing company was distributing a particular brand of soap through a large number of retail shops. Before a heavy advertisement campaign, the mean sales per week per shop was 140 dozens. After the campaign, a sample of 26 shops was taken and the mean sales was found to be 147 dozens with standard deviation 16. Can you consider the advertisement effective?

Sol.:

Given problem is a small sample mean test, as n = 26 which is less than 30. It is t-test

Mean, M = 140

Standard deviation, $\overline{S = 16}$, $\overline{X} = 147$

(i) Null Hypothesis, H_0: M = 140

Alternative Hypothesis, H_1 :M> 140 (Right tailed test)

(ii) Computing test statistic

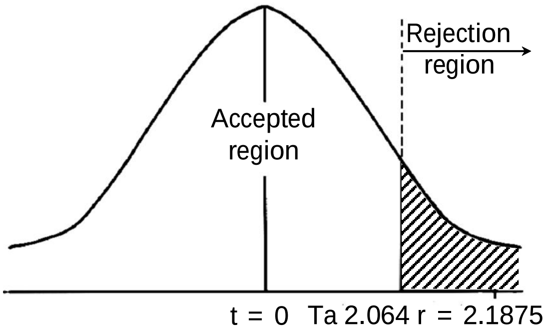
$$t = \frac{\overline{x} - M}{\frac{S}{\sqrt{n-1}}} = \frac{147 - 140}{\frac{16}{\sqrt{26-1}}} = \frac{7}{3.2} = 2.1875$$

(iii) Level of significate is $\alpha = 0.05$

(iv) Critical value

Degree of freedom = n – 1

= 26 – 1 = 25



t_α [0.05/25] = 1.708

(v) Deviation

Therefore, t_α lies in rejection region. Null hypothesis is rejected and alternative hypothesis is accepted at 5% level of significance i.e., advertisement campaign is effective.

35. In the past, a machine has produced washers having a thickness of 0.050 inches. To determine whether the machine is in proper working order, a sample of 10 washers is chosen for which the mean thickness is 0.053 and the standard deviation is 0.003 inches. Test the hypothesis at α = 0.05.

Ans :

Given,

Sample mean, $\overline{X}$ = 0.053

Population mean, μ_0 = 0.050

Population standard deviation, σ_t = 0.003

Sample size, n = 10

Let the null hypothesis be the average thickness of washers is of 0.050 inches.

$H_a : \mu_0 = 0.050$

$H_a : \mu_0 > 0.050$

The population is infinite, sample size is small ($\because$ n = 10) and population variance is not known. So the formula to be used is,

$$t = \frac{\bar{X} - \mu H_0}{\frac{\sigma_s}{\sqrt{n}}}$$

$$= \frac{0.053 - 0.050}{\frac{0.003}{\sqrt{10}}}$$

$$= 3.16$$

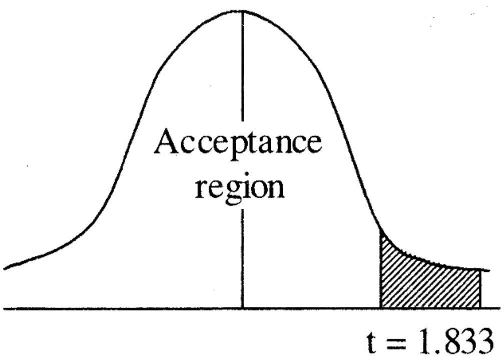
$$'t' = 3.16$$

Degree of freedom

$$d.f = n - 1$$

$$= 10 - 1$$

$$= 9$$



The rejection region for, d.f = 9

and α = 0.05 is t = 1.833

Right tailed test calculated t = 3.16

Since calculates value of 't' falls in the rejection region. H_0 is rejected and H_0 is accepted. We can conclude that the machine is in proper working condition.

36. A company is engaged in the manufacture of car tyres. Their mean life is 42,000 km with a standard deviation of 3000 km. A change in the production process is believed to improve the quality of tyres. A test sample of 28 new tyres has a mean life of 43,500 km. Do you think that the new car tyres are **significantly superior to the earlier one? Test the hypothesis at 5 percent level of significance.**

Ans :

Given that,

Sample size, η = 28

Sample mean, $\bar{x}$ = 43,500 km

Population standard deviation,

σ = 3000 km

Population mean, μ = 42,000 km

Null Hypothesis

$H_0 : \mu > 42{,}000$ km i.e., new car tyres are not superior to earlier car tyres.

Alternative Hypothesis

$H_0 : \mu = 42{,}000$ i.e., new car tyres are superior than the earlier car tyres.

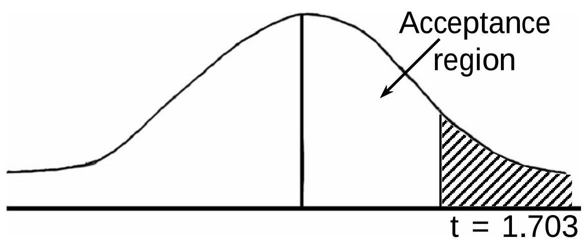
Test Statistic

$$t = \frac{\bar{x} - \mu}{\frac{\sigma_s}{\sqrt{n-1}}} = \frac{43{,}500 - 42{,}000}{\frac{3000}{\sqrt{28-1}}}$$

$$= \frac{1500 \times 5.196}{3000}$$

$$= \frac{7794}{3000} = 2.598$$

Degrees of freedom d.f (n – 1) = 27 at 5% level of j significance the tabulated value of $\boxed{t = 1.703}$.



Figure

Since, the calculated value of 't' does not lie in acceptance region, the null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted.

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

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

$$\text{We know that } S^2 = \frac{1}{n-1}\sum_i (X_i - \bar{X})^2 = \frac{1833.60}{9}$$

$\therefore$ Standard Deviation, $S = \sqrt{203.73} = 14.27$

(i) Null Hypothesis (H_0) : The data 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]

38. 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).

Sol : **(Aug.-21)**

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) : μ = 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

39. 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.

Sol : **(Oct.-20)**

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

$$= \bar{x} \pm t_{00.25} \frac{\sigma}{\sqrt{\eta}}$$

$$= 10.37 \pm 2.145 \; \frac{3.5}{\sqrt{15}}$$

$$= 10.37 \pm 1.93$$

$$= 10.37 + 1.93 = 12.31$$

$$= 10.37 - 1.94 = 8.43$$

Q67. Explain the test concerning the two means.

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

The more common situations involving tests on two means are those in which variances are unknown. If we assume that distributions are normal and that $\alpha_1 = \alpha_2 = \alpha$. The pooled t-test (often called the two-sample t-test) may be used. The test statistic is given by the following test procedure.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

With $n_1 + n_2 - 2$ degrees of freedom,

where

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2} \text{ or}$$

$$S_p^2 = \frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

Where $\bar{x}_1$, $\bar{x}_2$ are the means of two samples of size n_1 and n_2

S_1^2, S_2^2 are the variances of two samples of size n_1 and n_2.

The critical region wit this t-distribution can be obtained in a similar way.

For example when A.H. is $\mu_1 - \mu_2 \neq \delta$, the null hypothesis ($H_0$) is not rejected when

$-t_{\frac{\alpha}{2}, n_1+n_2-2} < t < t_{\frac{\alpha}{2}, n_1+n_2-2}$ and the critical region is $t < -t_{\frac{\alpha}{2}, n_1+n_2-2}$ or $t < t_{\frac{\alpha}{2}, n_1+n_2-2}$

Critical region for testing $H_0 : \mu_1 - \mu_2 = \delta$

Alternate hypothesis;		Reject null hypothesis if
(i)	$\mu_1 - \mu_2 \neq \delta$	$t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
(ii)	$\mu_1 - \mu_2 > \delta$	$t < t_\alpha$
(iii)	$\mu_4 - \mu_2 < \delta$	$t < -t_\alpha$

Note

1. The two-sample t-test can not be used if $\sigma_1 \neq \sigma_1$.
2. The two-sample t-test can not be used for "before and after" kind of data, where the data is naturally paired.

   In other words the samples must be "independent" for two sample t-test.

PROBLEMS

40. Two different methods of training the students in an institute gave the following results of marks.

Method A	**17**	**19**	**18**	**15**	**21**	**19**	**16**	**14**	**10**	**15**
Method B	**20**	**16**	**21**	**17**	**19**	**16**	**16**	**15**		

Test whether the two methods are significantly different at $\alpha = 0.05$.

Sol .:

Let the null hypothesis be that the two different methods are not significantly different.

$H_0 \quad : \mu_1 = \mu_2$

$H_a \quad : \mu_1 = \mu_2$

As the sample size is small and the population variance is not known t-distribution is used.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)\sigma_{s_2}^2 + (n_2-1)\sigma_{x_2}^2}{n_1+n_2-2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where, $$\sigma_{s_1}^2 = \frac{\Sigma(X_{1i} - A_i)^2 - [\Sigma(X_{1j} - A_1)^2 / n_1]}{n_1 - 1}$$

$$\sigma_{s_1}^2 = \frac{\Sigma(X_{2i} - A_2)^2 - [\Sigma(X_{2j} - A_2)^2 / n_2]}{n_1 - 1}$$

Given, $n_1 = 10; n_2 = 8$

Method A ($A_1 = 15$)				Method B ($A_2 = 19$)			
Sl.No.	X_{1i}	$(X_{1i} - A_1)$	$(X_{1i} - A_1)^2$	Sl.No.	X_{2i}	$(X_{2i} - A_2)$	$(X_{2i} - A_2)^2$
1	17	2	4	1	20	1	1
2	19	4	16	2	16	– 3	9
3	18	3	9	3	21	2	4
4	15	0	0	4	17	– 2	4
5	21	6	36	5	19	0	0
6	19	4	16	6	16	– 3	9
7	16	1	1	7	16	– 3	9
8	14	– 1	1	8	15	–4	16
9	10	–5	25				
10	15	0	0				
Total	**164**	**14**	**108**		**140**	**–12**	**52**

$$\bar{X}_1 = A_1 + \frac{\Sigma(X_{1i} - A_i)}{-n_1} = 15 + \frac{14}{10} = 16.4$$

$$\bar{X}_2 = A_2 + \frac{\Sigma(X_{2i} - A_2)}{n_2} = 19 + \frac{(-12)}{8} = 17.5$$

$$\sigma_{s_1}^2 = \frac{\Sigma(X_{1i} - A_1)^2 - [\Sigma(X_{2i} - A_2)^2 / n_2]}{n_1 - 1}$$

$$= \frac{108 - \frac{(14)^2}{10}}{10 - 1} = 9.82$$

$$\sigma_{s_1}^2 = \frac{\Sigma(X_{2i} - A_2)^2 - [\Sigma(X_{21} - A_2)^2 / n_1]}{n_2 - 1}$$

$$= \frac{52 - \frac{(-12)^2}{8}}{8 - 1} = 4.85$$

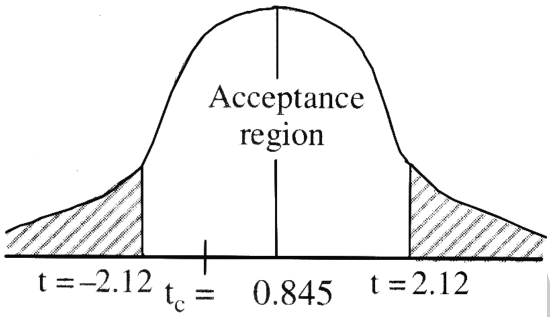
$$\therefore \ t = \frac{16.4 - 17.5}{\sqrt{\frac{(10-1)\,(9.82) + (8-1)(4.86)}{10 + 8 - 2}}\sqrt{\frac{1}{10} + \frac{1}{8}}}$$

$$= \frac{-1.1}{\sqrt{\frac{88.38+34.02}{16}}\sqrt{0.1+0.125}}$$

$$= \frac{-1.1}{2.77 \times 0.47} = -0.844$$

Degree of freedom df = $n_1 + n_2 - 2 = 10 + 8 - 2 = 16$

Rejection region for df = 16 at a = 0.05 is $|t| > 2.12$, for a two tailed test.



Calculate value of | t | = 0.84

As calculated value of $|t|$ falls in the acceptance region, H_0 is accepted. We can conclude that the two training methods are not significantly different.

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

(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 $n_1 = 10$, $n_2 = 10$

$$\therefore \quad \bar{X} = \frac{\Sigma X}{n} = \frac{55}{10} = 5.5 \; ; \; \bar{Y} = \frac{\Sigma Y}{n} = \frac{55}{10} = 5.5$$

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 $t = \dfrac{\bar{X} - \bar{Y}}{\sqrt{S^2\left[\dfrac{1}{n_1} + \dfrac{1}{n^2}\right]}} = \dfrac{5.5 - 5.5}{\sqrt{57.92\left(\dfrac{1}{10} + \dfrac{1}{10}\right)}}$

$$\frac{0}{3.403} = 0.$$

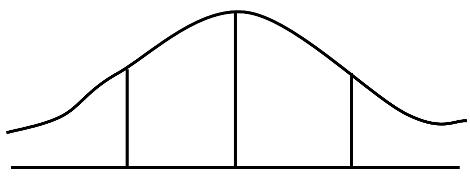
Degrees of freedom = $n_1 + n_2 - 2 = 10 + 10 - 2 = 18$

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



$t_{-\alpha}$ $H_1 = H_2$ t_α

3.6.3 Paired t-Test

Q68. Explain briefly about Paired t-Test.

(OR)

What is dependent sample (or) repeated measures t-test.

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

Testing of two means can be accomplished when data are in the form of paired observations. In this pairing structure, the conditions of the two populations are assigned randomly within homogeneous units.

Let $(X_n, X_{2l}), (X_{13}, X_{22}) \ldots (X_{ln}, X_{2n})$

be a set of *n* paired observations where we assume that the mean and variance of the population represented by X_1 are μ_1 and σ_1^2 and the mean and variance of the population represented by X_2 are μ_2 and σ_2^2.

Define the difference between each pair of observation as

$$D_i = X_{1i} - X_{2i} = 1 \text{ to } n$$

The D_1's are assumed to be normally distributed with mean and the variance σ_D^2.

$$\begin{aligned} \mu_D &= E(X_1 - X_2) \\ &= E(X_1) - E(X_2) \\ &= \mu_1 - \mu_2 \end{aligned}$$

Testing hypotehsis about $\mu_1 - \mu_2 = \delta$ can be accomplished by performing a one sample t-test on μ_D.

Specifically, testing H_0: $\mu_1 - \mu_2 = \delta$ against H_1: $\mu_1 - \mu_2 \neq \delta$ is equivalent to testing H_0: $\mu_D = \delta$ against H_1: $\mu_D \neq \delta$.

The test statistic for estig $\mu_D = \delta$ is,

$$t = \frac{\bar{D} - (\mu_1 - \mu_2)}{S_D / \sqrt{n}} = \frac{\bar{D} - \delta}{S_D / \sqrt{n}}$$

Which is a random variable with t-distribution, with n–1 degrees of freedom, where $\bar{D}$ is the sample mean and is given by, and S_D^2 is the sample variance and is given by,

$$\bar{D} = \frac{\sum_{i=1}^{n} D_1}{n}$$

$$S_D^2 = \frac{\left[\sum_{i=1}^{n} D_i^2 \frac{\left(\sum_{i=1}^{n} D_i\right)^2}{n}\right]}{n-1}$$

Confidence interval for $\mu_D = \mu_1 - \mu_2$ for Paired Observation

If $\bar{D}$ and S_D are the mean and standard deviation of the normally distributed differences of a random pairs of measurements, a$(1-\alpha)$ 100% confidence interval for $\mu_D = \mu_1 - \mu_2$ is,

$$\bar{D} - t_{\alpha/2}, \frac{S_D}{\sqrt{n}} < \mu_D < \bar{D} + t_{\alpha/2}, \frac{S_D}{\sqrt{n}}$$

Where $t_{\alpha/2}$, is the t-value with $v = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

PROBLEMS

42. The sales data of a product before and after special sales promotional campaign are given below.

Shop Efforts	Before Promotional Efforts	After Promotion
1	60	65
2	65	68
3	55	52
4	52	48
5	58	55
6	63	60
7	61	68
8	59	66

Can the promotional campaign be judged as effective?

Sol.:

Let the null hypothesis be that the promotional campaign has not been effective.

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1, < \mu_2$$

Alternate hypothesis = Promotional campaign has bee effective i.e., it has lead to an increase in sales.

Sample size small, population variance not known,

Paired t-test used.

$$t = \frac{\bar{d}}{s / \sqrt{n}}$$

$$\text{Where, } S = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n-1}} = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}} = \sqrt{\left(\frac{\Sigma d^2}{n-1}\right) - \frac{(\Sigma d)^2}{n(n-1)}}$$

Promotional Efforts			Difference	
Shop. No.	Before (1)	After (2)	d = (1) – (2)	d^2
1	60	65	–5	25
2	65	68	–3	9
3	55	52	3	9
4	52	48	4	16
5	58	55	3	9
6	63	60	3	9
7	61	68	–7	49
8	59	66	–7	49
			$\Sigma d = -9$	$\Sigma d^2 = 175$

$$\text{Mean, } \bar{d} = \frac{\Sigma d}{n} = \frac{-9}{8} = -1.125$$

$$S = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}} = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}}$$

$$= \sqrt{\frac{175 - 8(-1.125)^2}{8-1}}$$

$$= \sqrt{\frac{175 - 10.125}{7}}$$

$$= \sqrt{\frac{164.87}{7}} = \sqrt{23.55}$$

$$= 4.853$$

$$\therefore \quad S = 4.853$$

$$\therefore \quad t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{-1.125}{4.853/\sqrt{8}}$$

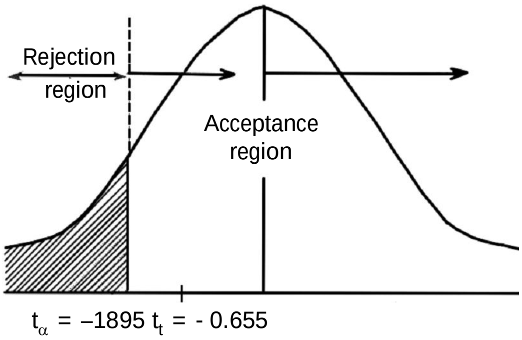
$$= \frac{-1.125}{1.715}$$

$$= -\ 0.6559$$

Decision

Critical value, d.f = n – 1 = 8 – 1 = 7

For 7 degrees of freedom, $t_{0.05} = 1.895$



Since, calculated value is less than table value, Null Hypothesis is accepted. Hence, it can be said that the promotional campaign has not been effective.

43. A certain stimulus administered to each of 12 patients results in the following changes in weight.

5, 2, 8, -1, 3, 0, -2,1, 5, 0, 4 and 6.

Can it be concluded that the stimulus will in general be accompanied by an increase in weight?

Sol.:

Let the null hypothesis be that there is no change in weight due to the stimulus.

$H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 < \mu_2$

(as we want to conclude that stimulus increases w)

As the sample size is small and population variance not known f-distribution is used. This is a case of related samples.

$$t = \frac{\bar{D}}{\frac{\sigma_{diff}}{\sqrt{n}}}$$

where, $\sigma_{diff} = \sqrt{\frac{\Sigma D^2 - (\bar{D})^2.n}{n-1}}$

In this question, the changes i.e., difference is given

Difference (D)	5	2	8	-1	3	0	-2	1	5	0	4	6
D^2	25	4	64	1	9	0	4	1	25	0	16	36

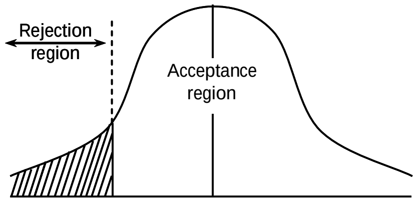
$$\bar{D} = \frac{\Sigma D}{n} = \frac{31}{12} = 2.58$$

$$\sigma_{diff} = \sqrt{\frac{\Sigma D^2 - (\bar{D})^2.n}{n-1}} = \sqrt{\frac{185 - 12(2.58)^2}{12-1}} = \sqrt{\frac{105.213}{11}} = 3.09$$

$$r = \frac{\bar{D}}{\frac{\sigma_{idff}}{\sqrt{n}}} = \frac{2.58}{\frac{3.09}{\sqrt{12}}} = \frac{2.58}{0.892} = 2.8923$$

Degrees of freedom, d.f = n – 1 = 12 – 1 = 11

Rejection region for d.f = 11 and α = 0.05 is $|t| < -2.201$, for left tail test.



Calculated value of t = 2.8923 [$\therefore t_{calculated} > t_{tabulated} \rightarrow$ reject H_0]

Since, calculated value falls in the rejection region, H_0 is rejected and H_a is accepted. We can conclude that the stimulus is effective in increasing the weight.

44. A company has reorganized its sales department. The following data shows its weekly sales both before and after reorganization.

Week No.	1	2	3	4	5	6	7	8
Sales prior to reorganization	**12**	**15**	**13**	**11**	**17**	**15**	**10**	**11**
Sales after reorganization	**16**	**17**	**14**	**13**	**15**	**14**	**12**	**11**

Can it be concluded that the reorganization of the sales department of the company has resulted in a significant increase in its sales?

Sol :

Let the null hypothesis be the average sales before and after reorganization are same.

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$ (as we want to conclude that reorganization increase sales)

As the sample size is small and population variance not known student *t* distribution is used. This is a case of related samples.

Week No.	Before Reorganization	After Reorganization	Difference (D)	D^2
1	12	16	– 4	16
2	15	17	– 2	4
3	13	14	– 1	1
4	11	13	– 2	4
5	17	15	2	4
6	15	14	1	1
7	10	12	– 2	4
8	11	11	0	0
		Total	**– 8**	**34**

Test statistic under H_0, in the usual notations, the appropriate test statistic is,

$$t = \frac{\bar{D}}{\sigma/\sqrt{n}} \sim t_{n-1} = t_{8-1} = t7$$

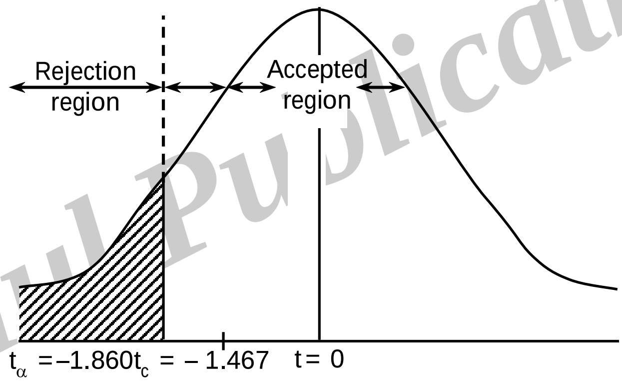
$$\bar{D} = \frac{\Sigma D}{n} = \frac{-8}{8} = -1.0$$

$$\sigma^2 = \sqrt{\frac{1}{n-1}\left[\Sigma D^2 \frac{(\Sigma D)^2}{n}\right]} \Rightarrow \sqrt{\left[\frac{\Sigma D^2 n(\bar{D})^2}{n-1}\right]}$$

$$= \sqrt{\frac{34-8(-1)^2}{8-1}} \Rightarrow \sqrt{3.714} = 1.927$$

$$\therefore\ t_7 = \frac{\bar{D}}{\sigma/\sqrt{n}} = \frac{-1}{\frac{1.927}{\sqrt{8}}} \Rightarrow \frac{-1}{0.6813} = -1.467$$

Tabulated value of t at 5% level of significant and for 8 d.f. = 1.860.



Since calculated $|t|$ = 1.467, which is less than tabulate t, it is not significant at 5% level of significance. Hence, H_0 is not rejected at 5% level of significance and we conclude that the reorganisation is ineffective in increasing sales *i.e.*, the sales are same both before and after reorganisation.

3.7 Hypothesis Formulation and Testing

Q69. Define Hypothesis. Explain the procedure for testing a statistical hypothesis.

Ans :

Meaning

Statistical hypothesis is an assumption about the parameters of the population and sometimes it also concerns the type and nature of the distribution.

Example

(i) The average height of soldiers in the army is 165 cm.

(ii) A given drug cures 80% of the patients taking it.

(iii) A given machine has an effective life of 20 years.

All these hypotheses may be verified on the basis of certain sample tests. Procedures or tests which enable us to decide whether to accept or reject the hypothesis are called tests of hypothesis or tests of significance.

Use

The test of hypothesis discloses the fact whether the difference between the computed statistic and hypothesical parameter is significant or otherwise.

Procedure

Hypothesis Testing

To test a hypothesis means to tell on the basis of the data, whether or'not the hypothesis seems to be valid. Procedure for hypothesis testing refers to all those steps that we undertake for making a choice between two actions i.e., rejection and acceptance of a null hypothesis.

Steps in Hypothesis Testing

1. Making a Formal Statement

This- step consists of making a formal statement of the null hypothesis *(H_o)* and also of alternatively hypothesis (H_0). This means that the hypothesis should be clearly stated, considering the nature of the research problem.

Example

In a population of 30 industrial units of same size, labour turnover ig analyzed. The past records show that mean of turnover is 320 employees. A sample of 5 of these units is taken at random which gave a mean of 300. The labour minister wants to know if there is a significant difference between the sample and population.

Null hypothesis $H_0 : \mu = 320$

Alternative hypothesis $H_0 : \mu \neq 320$

In the above example, if the union claims that the turnover is above 320, and if the claim has to tested then

Null hypothesis $H_0 : \mu = 320$

Alternative hypothesis $H_0 : \mu = 320$

We should decide whether to use a two tailed test or one-tailed test.

2. Selecting a Significance Level

The hypothesis are tested on a predetermined level of significance and as such the same should be specified. Generally, in practice, either 5% level or 1% is adopted for the purpose.

The factors that affect the level of significance are,

- The size of the samples
- Magnitude of the difference between sample means.
- Variability of measurements within samples.
- Whether the hypothesis is directional or non-directional.

The level of significance must be adequate in the context of the purpose and nature of inquiry.

3. Deciding the Distribution to Use

The next step is to determine the appropriate sampling : distribution. The choice generally remains between normal distribution and the t-distribution.

4. Selecting a Random Sample and Computing an Appropriate Value

Another step is to select a random sample (s) and compute an appropriate value from the sample data concerning the test statistic utilizing the relevant distribution.

5. Calculation of Probability

Calculate the probability that the sample result would diverge as widely as it has from expectations, if null hypothesis were is fact true.

6. Comparing the Probability

Now, compare, the probability that has been calculated with the specified value for a, the level of significance. If the calculated probability is less than or equal to a value in case of one tailed test and — value in case of two tailed test, then reject H_o and accept H_a and vice versa.

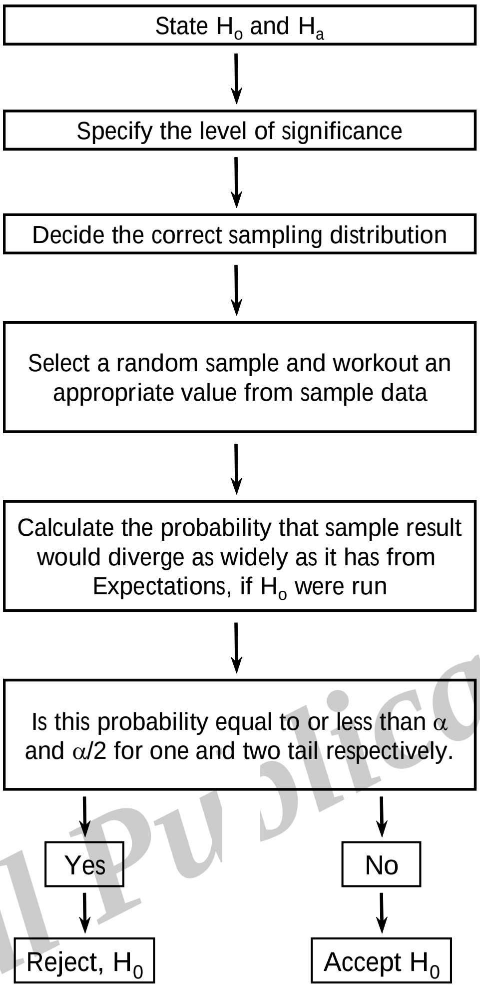


Fig.: Steps in Hypothesis

Q70. Explain different types of hypothesis tests.

(OR)

Classify the different types of hypothesis test.

Ans : **(Imp.)**

1. Question Hypothesis

It is regarded as the simplest form of empirical observation. Under this type, the hypothesis is usually represented in a simple question form. However, it does not serve as the definition of hypothesis i.e., it does not convey the meaning of hypothesis. This type of hypothesis is mostly used in those research projects which needs to find out an answer to a simple question.

Example

Is there any effect of reinforced advertisement on the product sales volume?

2. Declaration Statement

It declaration hypothesis, the researcher basically builds/develops an imaginary relationship between two independent variables. On the basis of study by using the historical data.

3. Directional Hypothesis

Directional hypothesis, basically emphasizes upon drawing an expected direction for the existing relationship between different variables.

4. Null/Non-Directional Hypothesis

The non-directional hypothesis is explained in the null (Negative) form i.e., it states that there exists no relationship between the two different variables. This type of hypothesis is statistical in nature and can be measured within the framework of the probability theory. It usually gives a negative outcome to the researcher. Therefore, it is considered and called as non-directional in nature.

Q71. Define the following terms :

(i) Null hypothesis

(ii) Alternate hypothesis

(iii) Simple hypothesis

(iv) Composite hypothesis

Ans :

i) Null Hypothesis

A definite statement about the populations parameter for applying the tests of significance is called the null/hypothesis, which is usually a hypothesis of an difference.

For example if we want to decide whether one procedure is better than another, we formulate the null hypothesis that there is no difference between the procedure. It is usually denoted by H_0.

Definition

According to Prof. R.A. Fisher. "Null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true".

Example : $H_o\mu = \mu H_0$

ii) Alternate Hypothesis

To be able to construct suitable criteria for testing statistical hypothesis, it is necessary that we also formulate alternative hypothesis. Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis and is usually denoted by the symbol H_y For example if we want, to test the null hypothesis that the average height of the soldiers is 165 cm

Example: $H_0: \mu \neq \mu_{H0}, \mu > \mu_{H0}, \mu < \mu_{H0}$

iii) Simple Hypothesis and Composite Hypothesis Simple Hypothesis

Simple hypothesis is a statistical hypothesis which completely specifies an exact parameter *p.*

Null hypothesis is always a simple hypothesis stated as an equality specifying an exact value of the parameter.

Examples

1. HQ . P – ftp

   i.e., population mean equals to a specified constant p_0.

2. F : $P_t - P_2 = 8$

   i.e., the difference between the sample means equal to a constant 8.

iv) Composite Hypothesis

Composite hypothesis is stated in terms of several possible values i.e., by an inequality.

Alternative hypothesis is a composite hypothesis involving statements expressed as inequalities such as <, > or $\neq$.

Example

1. $H_1 : \mu > \mu_0$
2. $H_1 : \mu < \mu_0$
3. $H_1 : \mu \neq \mu_0$

Q72. Explain various errors in hypothesis.

Ans :

The main objective in sampling theory is to draw valid inferences about the population parameters on the basis of the sample results. There is always a chance of making error. There are two possible types of errors in the test of hypothesis.

(i) Type I error

(ii) Type II error

(i) Type I Error

If we reject the null hypothesis when it should be accepted, we say that a type I error has been made. The probability of committing a type I error is denoted by a i.e,

P[Reject H_0 when it is true] = α

(ii) Type II Error

If we accept the null hypothesis when it should be rejected, we say that a type II error has been made. The probability of committing a type II error is denoted by β i.e.,

P[Accept H_0 when it is wrong] = β

Actual	**Decision**	
	Accept H_0	**Reject H_0**
H_0 is true	Correct decision (No error) probability = $1 - \alpha$	Wrong decision (Type I error) probability = α
H_0 is wrong	Wrong decision (Type II error) probability = β	Correct decision (No error) probability $1 - \beta$

Table: Dichotomous

α error is also known as producer's risk because a good lot is getting rejected.

β error is also known as consumer's risk because a bad lot is getting accepted.

Trade off Between Type I and Type II Error

Type I error can be controlled by fixing it at a lower significance level (say 1% instead of 5%). But with a fixed sample size n, when we try to reduce type I error, the probability of committing type II error increases. Both types of errors cannot be reduced simultaneously. Hence here is a trade off between these two types of errors.

Q73. Define the following terms :

(i) Level of significance

(ii) One tailed and Two tailed test

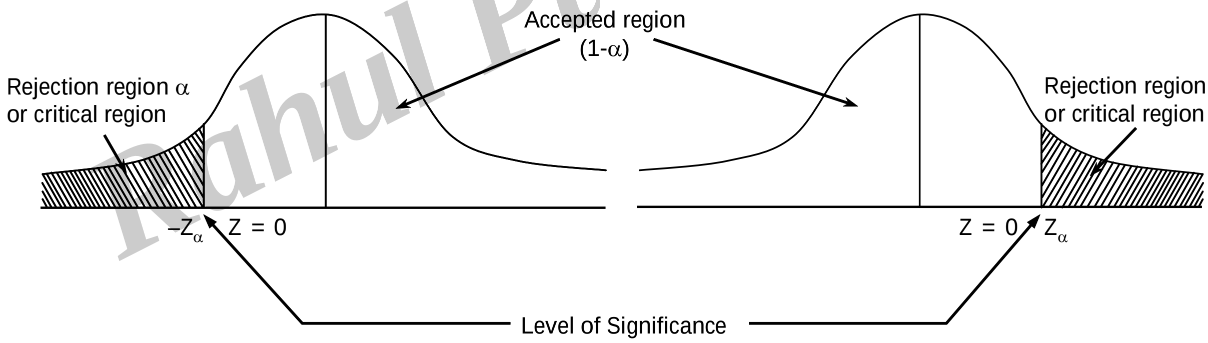
Ans :

(i) Level of significance

In testing of a given hypothesis the maximum probability with which we would be willing to risk a type I error is called the level of significance of the test.

It is denoted by 'a' and in other words it is the maximum probability of making type I error, a is generally taken as 0.05 or 0.01 (i.e, 5% and 1%). α = 0.01 is used for high precision and α = 0.05 for moderate precision.

If we adopt 5% level of significance, it implies that there are 5 chance in 100, we are likely to reject a correct null hypothesis H_0. In other words, we are 95% confident that we have made a correct decision. The level of significance is also called the critical level.



(a) Power of a Test

Power of a test is the probability of rejecting H_0 given that a specific alternative is true. The power of a test can be computed as $1 - \beta$.

Often different types of tests are compared by contrasting power properties.

(b) Critical Region (C.R)

The area under the probability curve is divided into two regions.

(i) Region of rejection (significant region) where N.H.

(ii) Region of acceptance (nonsignificant region) where N.H. is accepted.

Critical region is the region of rejection of N.H. The area of the critical region equals to the level of significance α.

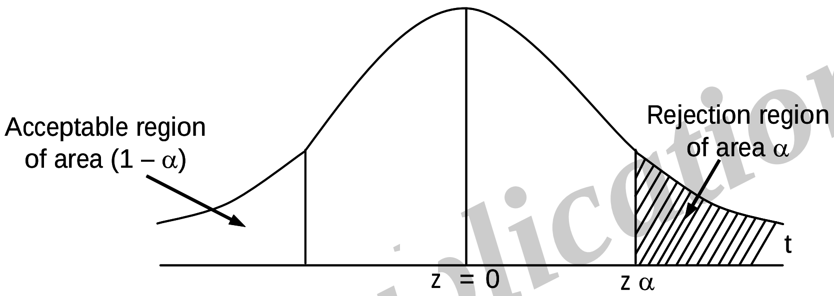
(ii) One Tailed and Two Tailed Tests

A test of any statistical hypothesis where the alternative hypothesis is expressed by the symbol (<) or the symbol (>) is called a cone tailed test since the entire critical region lies in one tail of the distribution of the test statistic.

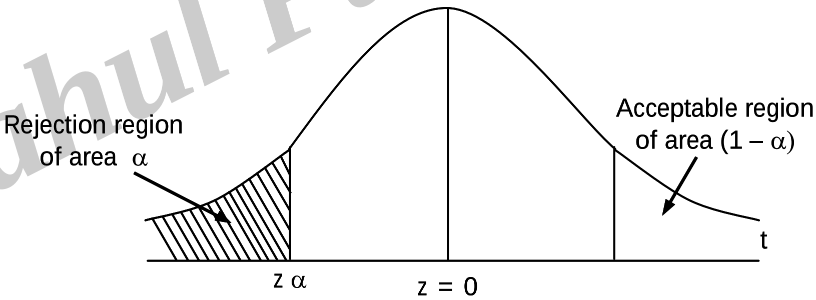
The critical region for all alternative hypothesis containing the symbol (>) lies entirely in the right tail of the distribution while the critical region for an alternative hypothesis containing the symbol (>) lies entirely in the left tail. The symbol indicates the direction where the critical region lies.

A test.of any statistical hypothesis where the alternative is written with a symbol ($\neq$) is called a two tailed test, since region is split into two equal parts, one in each tail of the distribution of the test statistic.

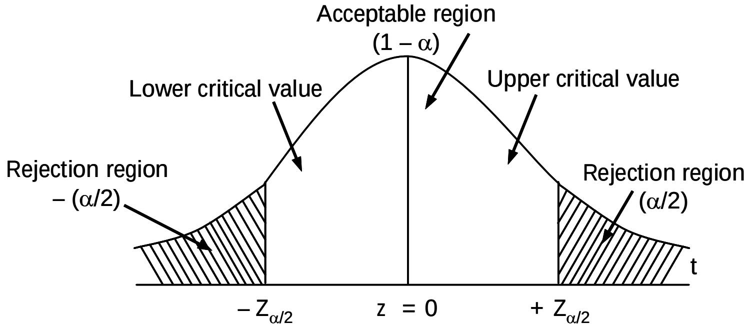
Right Tailed Test



Left Tailed Test



Two Tailed Test



Short Question & Answers

1. Define Classification.

Ans :

Classification is the grouping of related facts into classes.

i) **According to Connor (1997)** "The process of arranging things in groups or classes according to their resemblances and affinities and gives expression to the unity of attributes that may subsist amongst a diversity of individuals". "Classification is the process of arranging data into sequences and groups according to their common characteristics or separating them into different but related parts".

ii) **According to Secristi** "The process of grouping large number of individual facts and observations on the basis of similarity among the items is called classification".

2. Define tabulation. State the objectives of tabulation.

Ans :

The simplest and most revealing devices for summarizing data and presenting them in a meaningful manner is the statistical table. After classifying the statistical data, next step is to present them in the form of tables. A table is a systematic organization of statistical data in rows and columns. The purpose of a table is to simplify the presentation and to facilitate comparisons. The main objective of tabulation is to answer various queries concerning the investigation. Tables are very helpful for doing analysis and drawing inferences from them.

Classification and tabulation go together, classification being the first step in tabulation. Before the data are put in tabular form, they have to be classified.

3. Diagrammatic Representation of Data

Ans :

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).

4. Diagrams

Ans :

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.

5. Graphical representation of data

Ans :

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.

6. Differentiate between diagrams and graphs.

Ans :

Sl.No.	Diagrams	Sl.No.	Graphs
1.	Diagrams are usually constructed on a plain sheet and they are use ful 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 different types of diagrams.	3.	Graphs provide reliable, accurate and authentic information.

7. Define Statistics.

Ans :

Introduction

The. term 'Statistics' has been defined differently by different authors. Some have defined the word in the sense of numerical data, whereas others have defined in the sense of statistical methods.

A few definitions have been discussed here.

i) According to Bowel defined Statistics as 'Numerical statement of facts in any department of enquiry placed in relation to each other" According to Yule and Kendall "By statistics we mean quantitative data affected to a marked extent by multiplicity of causes'.

ii) A comprehensive definition is given by Prof. Horace Secrets that points out all characteristics that numerical data must possess to be called statistics.

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

8. Functions of statistics.

Ans :

Following are the important functions of Statistics

1. **Systematic Collection and Presentation of Facts** : Facts backed by objective numbers carry lot more conviction. Statistics ensures that data is' collected in a systematic manner and presented without any subjective bias. Statistics adds a bit of precision and definiteness to general statements, resulting in greater conviction.

2. **Simplification of Mass Figures** : Statistics cuts through the clutter of data normally available and presents the facts effectively. The large mass of data collected is transferred into a few critical figures, which help in overall analysis and interpretation.

3. **Facilitates Comparison :** Statistics enables impartial and fair comparison of data across a wide range of competing alternatives. For example, Inflation rate compares the current prices with prices of previous year and indicates the trend of rising prices. Thus, Statistics performs the function of providing a definite meaning to the state of affairs by quantification of data and comparison.

4. **Helps in Formulation and Testing of Hypothesis :** Statistics has developed into a separate discipline, with its own theories, principles and methodology, which can be used to answer specific queries. For example, if a contractor has to decide between two alternative choices of cement, he has to check whether the quality of cement is satisfactory. He can take the help of statistics and use sampling techniques and principles to arrive at a decision. Similarly, Statistics can be used in determining the result of a medicine on a patient, the impact of newly introduced Government policy on the general public, and whether the purchase of a computer has resulted in efficiency gains etc.

9. Define mode.

Ans :

Meaning

Mode may be defined as the value that occurs most frequently in a statistical distribution or it is defined as that exact value in the ungrouped data if each sample which occurs most frequently.

Definitions

(i) **According to Croxton and Cowden,** "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values."

(ii) **According to A.M. Tuttle,** "Mode is the value which has the greatest frequency density in its immediate neighbourhood."

(iii) **According to Zizek,** "The mode is the value occurring most frequently in a series of items and around which the other items are distributed most densely."

10. Define Dispersion.

Ans :

Meaning

The concept of dispersion is related to the extent of scatter or variability in observations. The variability, in an observation, is often measured as its deviation from a central value. A suitable average of all such deviations is called the measure of dispersion. Since most of the measures of dispersion are based on the average of deviations of observations from an average, they are also known as the averages of second order.

Definitions

As opposed to this, the measures of central tendency are known as the averages of first order. Some important definitions of dispersion are given below:

(i) **According to A.L. Bowley,** "Dispersion is the measure of variation of the items."

(ii) **According to Connor,** "Dispersion is the measure of extent to which individual items vary."

(iii) **According to Simpson and Kafka,** "The measure of the scatteredness of the mass of figures in a series about an average is called the measure of variation or dispersion."

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

11. Define Coefficient of Variation.

Ans :

Coefficient of Variation (CV) was proposed by Karl Pearson. It is used to compare the variability of two (or) more distributions. A distribution with greater C.V. is considered as more variable or less consistent, less unit form, less stable or less homogeneous distribution and the distribution with less C.V is considered as less variable or more consistent, more uniform, more stable or more homogeneous distribution.

Coefficient of variation =

$$\frac{\text{Standard deviation}}{\text{Arithmetic mean}} \times 100$$

12. When is t-test used?

Ans :

Meaning

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.

13. Testing for One and Two Means

Ans :

To test if sample mean $\overline{x}$ differs significantly from the hypothetical value m of the population mean, we use the following t-test.

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

Where,

n = Sample Size

$n = \frac{1}{n} \sum x_1$ (Sample mean)

μ = Population mean from which sample is taken

$s = \sqrt{\frac{1}{n-1}\Sigma(x_1 - \overline{x})^2}$ (S.D. Sample)

Degree of freedom = n – 1

Decision

If calculate value of $|t| \geq$ table value for (n–1) d.f. at 5% level then reject the hypothesis and say atht the two means differ significantly.

The sample does not seem to have been drawn from the population with mean μ.

If calcualted value of $|t| <$ table value for (n – 1) d.f. at 5% level then accept the hypothesis.

14. Explain the test concerning the two means.

Ans :

The more common situations involving tests on two meansiyare those in which variances are unknown. If we assume that distributions are normal and that $\alpha_1 = \alpha_2 = \alpha$. The pooled t-test (often called the two-sample t-test) may be used. The test statistic is given by the following test procedure.

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - \delta}{S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

With $n_1 + n_2 - 2$ degrees of freedom,

where

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2} \text{ or}$$

$$S_p^2 = \frac{\Sigma(x_i - \overline{x})^2 + \Sigma(y_i - \overline{y})^2}{n_1 + n_2 - 2}$$

Where $\overline{x}_1$, $\overline{x}_2$ are the means of two samples of size n_1 and n_2

S_1^2, S_2^2 are the variances of two samples of size n_1 and n_2.

The critical region wit this t-distribution can be obtained in a similar way.

For example when A.H. is $\mu_1 - \mu_2 \neq \delta$, the null bypotehsis ($H_0$) is not rejected when

$-t_{\frac{\alpha}{2}, n_1+n_2-2} < t < t_{\frac{\alpha}{2}, n_1+n_2-2}$ and the critical region is $t < -t_{\frac{\alpha}{2}, n_1+n_2-2}$ or $t < t_{\frac{\alpha}{2}, n_1+n_2-2}$

Critical region for testign $H_0 : \mu_1 - \mu_2 = \delta$

Alternate hypothesis;	Reject null hypothesis if
(i) $\mu_1 - \mu_2 \neq \delta$	$t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
(ii) $\mu_1 - \mu_2 > \delta$	$t < t_{\alpha}$
(iii) $\mu_4 - \mu_2 < \delta$	$t < -t_{\xi}$

15. Paired t-Test.

Ans :

Testing of two means can be accomplished when data are in the form of paired observations. In this pairing structure, the conditions of the two populations are assigned randomly within homogeneous units.

Exercise Problmes

1. Draw a less than ogive from the following and locate the median:

Size:	10–20	20-30	30-40	40-50	50-60
Frequency :	20	60	100	150	75

[**Ans.:** 41.5]

2. Calculate the value of mean for the following data:

Mid point:	10	15	20	25	30	35	40
Frequency :	7	9	18	26	10	4	3

[**Ans.:** $\bar{X}$ = 23.05]

3. Calculate mean, median and mode from the following frequency distribution of marks at a test in English.

Marks:	5	10	15	20	25	30	35	40	45	50
No. of Students:	20	43	75	76	72	45	39	9	8	6

[**Ans.:** $\bar{X}$ = 22.16, Med. = 20, Mo. = 20]

4. A random sample of 16 values from a normal population showed a mean of 48.5 and sum of squres of deviations from the mean equal to 135. Can it be ssumed that the mean of the population is 43.5? (Use 5 per cent level of significance.)

[**Ans.:** t = 2.67]

5. A random sample of 12 paris of observations from anormal population gives a coefficeinto f crrelation of 0.54. Is this value significant of correlation in the population?

[**Ans.:** t = 1.597]

6. In an examination in Psychology, 12 students in one class had a mean grade of 78 with a standard deviation of 6, while 15 students in another class had a mean grade of 74 with a standard deviation of 8. Is there a significant difference between the means of the two groups?

[**Ans.:** t= 1-441]

7. In a random sample of 600 males in Jaipur, 400 were found to be smokers while in another random sample of 900 females in Delhi, 450 were found to be smokers. Discuss the question whether the data reveal a significant difference in Jaipur to Delhi so far as the proportion of smokers is concerned.

[**Ans.:** Diff./S.E. 6-421]

Choose the Correct Answer

1. What is the arrangement of data in rows and columns known as? [c]
   (a) Frequency distribution (b) Cumulative frequency distribution
   (c) Tabulation (d) Classification

2. When the quantitative and qualitative data are arranged according to a single feature, what is the tabulation known as? [a]
   (a) One-way (b) Bivariate
   (c) Manifold division (d) Dichotomy

3. Which function does the tabulation origin spot specify? [a]
   (a) The list of integers (b) The list of maxterms
   (c) The list of minterms (d) None of the above

4. What does the tabulation form exercise? [c]
   (a) Gates (b) Demorgan's postulate
   (c) Matching process cycle (d) Venn diagram

5. What was the first tabulation method known as? [d]
   (a) Quine-McCluskey (b) Cluskey
   (c) McQuine (d) None of the above

6. What is the table where the variables are subdivided with interrelated features known as? [d]
   (a) Order level table (b) Sub-parts of a table
   (c) One-way table (d) Two-way table

7. In a tabular presentation, what is the summary and presentation of data with different non-overlapping classes defined as? [a]
   (a) Frequency distribution (b) Chronological distribution
   (c) Ordinal distribution (d) Nominal distribution

8. What are the general tables of data used to show data in an orderly manner known as? [c]
   (a) Double characteristic tables (b) Manifold tables
   (c) Repository tables (d) Single characteristics tables

9. The runs scored by a batsman in 5 ODIs are 31,97,112, 63, and 12. The standard deviation is [c]
   (a) 24.79 (b) 23.79
   (c) 25.79 (c) 26.79

10. Find the mode of the call received on 7 consecutive day 11,13,13,17,19,23,25 [b]
    (a) 11 (b) 13
    (c) 17 (d) 23

Fill in the blanks

1. The ________ is defined as the frequency distribution graph of a series.
2. Every ________ must be given a suitable title.
3. The ________ selected must satisfy all the values to be plotted on the graph.
4. A frequency ________ is a graph representing frequency distribution.
5. One of the most commonly used and easily understood methods for graphical represen- tation of frequency distribution data is called ________.
6. A smoothed ________ 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.
7. A ________ Diagram is a type of graph that displays data in a circular graph.
8. ________ of the data is a necessary function of any statistical analysis.
9. ________ is the simplest method of studying dispersion.
10. The standard deviation concept was introduced by Karl Pearson in ________ .

Answers

1. Ogive
2. Diagram
3. Scale
4. Polygon
5. Histogram
6. Frequency
7. Pie
8. Summarisation
9. Range
10. 1823

One Mark Answers

1. Geographical Classification

Ans :

When data are classified with reference to geographical locations such as countries, states, cities, districts etc, it is known as geographical classification.

2. What is frequency distribution?

Ans :

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.

3. Simple Bar Diagram

Ans :

A simple bar diagram is used to represent only one variable. It should be kept in mind that, only length is taken into account and not width.

4. What is Range ?

Ans :

Range is the simplest method of studying dispersion. It is defined as the difference between the value of the smallest item and the value of the largest item included in the distribution.

5. Mean deviation.

Ans :

"Mean Deviation of a series is the arithmetic average of the deviations of various items from a measure of central tendency.

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.

4.1 Analysis of Variance

Q1. What is an ANOVA? State the assumptions and applications of an ANOVA.

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

Meaning

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 quantiative population 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 of wheat and yield of wheat and yield of wheat on these plots is given. We are interested in finding out whether the samples have come from the same population. ANOVA answers this question.

Assumptions

Analysis of variance test is based on the test statistic F (or variance ratio).

It is based on the following assumptions.

(i) Observation 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

The applications of ANOVA are as follows,

1. Anova is used in education, industry, business, psychology field 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)**

The following are the various steps under One way Anova :

1. Set the null hypothesis H_0 and alternate hypothesis H_1.
2. Computation of Test statistic. The test statistic is obtained by the following steps :

Steps I :

Find the sum of the values of all the items of all the samples and denote it by T, i.e., obtain

$$T = \sum_i \sum_j x_{ij}.$$

Step II :

Compute the correction factor C, where

$C = \frac{T^2}{N}$ .N is the total number of all the observations of all the samples and T is obtained in STEP 1.

Step III :

Find the sum of the squares of all items of all the samples i.e., find $\sum_i \sum_j x_{ij}^2$.

Step IV :

Find the total sum of the squares SST, where

$$SST = \sum_i \sum_j x_{ij}^2 - C, \text{ or } SST = \sum_i \sum_j x_{ij}^2 - \frac{T^2}{N}.$$

where C is the correction factor obtained in STEP 2.

Step V :

Find the sum of the squares between the samples, i.e., SSC by the following steps

(i) Square the totals and divide it by the number of items in each sample.

$$P = \frac{(\Sigma X_i)^2}{n_i} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3}$$

(ii) Subtract the correction factor $C = \frac{T^2}{N}$ from P, then the resulting figure would be the sum of the squares between the samples SSC, i.e.,

$$SSC = P - \frac{T^2}{N}$$

Step VI :

Find the sum of the squares within the samples SSE from the result

SSE = SST – SSC.

[From Step 4 and Step 5]

Step VII :

Set up the ANOVA table and calculate F, which is the test statistic.

3) Level of Significance. Let the level of significance be $\propto = 0.05$.

4) Decision

(a) If the computed value of F is less than the tabled value of F at $\alpha = 0.05$, then accept the null hypothesis H_0.

(b) If the computed value of F is greater than the tabled value of Fat $\alpha = 0.05$, then reject the null hypothesis H_0 and accept the alternate hypothesis H_1.

PROBLEMS

1. A farmer applied three types of fertilizers on 4 separate plots. The figure on yield per acre are tabulated below:

Fertilizers Plo 2 ts	Yield				
	A	B	C	D	Total
Nitrogen	6	4	8	6	24
Potash	7	6	6	9	28
Phosphates	8	5	10	9	32

Find out if the plots are materials different in fertility, as also, if the three fertilizers make materials difference in yields.

Sol :

Correction Factor : $C = \frac{T^2}{N} = \frac{(84)^2}{12} = 588.$

SST = Total sum of square = $\Sigma\Sigma x_{ij}^2 - C$

$= [(6^2 + 7^2 + 8^2) + (4^2 + 6^2 + 5^2) + (8^2 + 6^2 + 10^2) + (6^2 + 9^2 + 9^2)] - 588$

$(36 + 49 + 64) + (16 + 36 + 25) + (64 + 36 + 100) + (36 + 81) + (81) - 588$

$= 624 - 588 = 36$

SSC = sum of square between plots $= \frac{\Sigma\left(\sum_j x_i\right)^2}{r} - C$

$= \left[\frac{(21)^2}{3} + \frac{(15)^2}{3} + \frac{(24)^2}{3} + \frac{(24)^2}{3}\right] - 588 = \frac{1818}{3} - 588$

$= \frac{441}{3} + \frac{225}{3} + \frac{576}{3} + \frac{576}{3} - 558$

$= 147 + 75 + 192 + 192 - 568$

$= 606 - 558 = I8$

Sum of squares between Fertilizers $= \left(\sum_{i=1}^{r} T_i^2 / c\right) - C$

$= \frac{(24)^2}{4} + \frac{(28)^2}{4} + \frac{(32)^2}{4} - 588$

$= \frac{576}{4} + \frac{784}{4} + \frac{1024}{4} - 588$

$= \frac{2384}{4} - 588 = 596 - 588 = 8$

Error sum of squares = 36 – 18 – 8 = 10

Null Hypothesis : (i) Plots are equally fertile

(ii) Fertilizer are equally effective

ANOVA Table

Prices of variation	**d.o.f.**	**S.S.**	**$MM = \frac{S.S.}{d.o.f}$**	**FC**	**F_{Tab}(5%)**
Between Plots	3	18	$MSC = \frac{18}{3} = 6$	$\frac{MSC}{MSE} = 3.6$	4.76
Between Fertilizer	2	8	$MSR = \frac{8}{2} = 4$	$\frac{MSR}{MSE} = 2.4$	5.14
Error	6	10	$MSE = \frac{10}{6} = 1.667$		
Total	**C_{r-1} = 11**	**36**			

Conclusion

We accept the hypothesis at (i) and (ii), what is, the plots are equally effective and the fertilizer have the same effect.

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

(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
$T_1 \rightarrow 35$	$T_2 \rightarrow 25$	$T_3 \rightarrow 30$

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

$$= \frac{\Sigma T_j^2}{n_j} - C.F$$

$$= \left[\frac{(35)^2}{5} + \frac{(25)^2}{5} + \frac{(30)^2}{5}\right] - 540$$

$$= \frac{1225}{5} + \frac{625}{5} + \frac{900}{5} - 540$$

= 245 + 125 + 180 – 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	$26/12 = 2.166$	$F_{cal} = \frac{5}{2.166} = 2.308$
TSS	15 – 1 = 14	36	–	

Conclusion :

F_{cal} & F_{Ratio} = 2.166

F_{tab} at (2, 12) degree of freedom F_{tab} = 3.74

$F_{tab} < F_{tab}$

∴ H_0 is accepted

We conclude that there is no difference in the teaching method.

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

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)

$$\text{SSC} = \sum_i \frac{T_j^2}{n_j} = \frac{(GT)^2}{N}$$

$$\left[\frac{(205)^2}{5}+\frac{(204)^4}{5}+\frac{(248)^2}{5}+\frac{(193)^2}{5}\right] - 36{,}125$$

= 8405 + 8323.2 + 12300.8 + 7449.8 – 36125

= 36528.8 – 36125

= 403.8

Sum of Squares between Rows

$$= \frac{(173)^2}{4} + \frac{(189)^2}{4} + \frac{(154)^2}{4} + \frac{(158)^2}{4} + \frac{(176)^2}{4} - \text{CF}$$

$\Rightarrow$ 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	23.7 629	(c – 1) (r – 1) 3 × 4 = 12 19	$\frac{23.7}{12}$ = 1.975	—

$F_{0.05}$ (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.

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 Classified 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 facilitated 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

4. 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	Drugs				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)

$$(\text{TSS}) = \sum_i \sum_j X_{ij}^2 - \text{C.F}$$

$$= [(19)^2 + (8)^2 + (23)^2 + (8)^2 + (10)^2 + (9)^2 + (12)^2 + (6)^2 + (11)^2 + (10)^2 + (13)^2 + (16)^2] - \text{C.F}$$

$$= 361 + 64 + 529 + 64 + 100 + 81 + 144 + 36 + 121 + 100 + 169 + 256 - 1752.08$$

$$= 2025 - 1752.08$$

$$\therefore \quad \text{TSS} = 272.92$$

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)

$$\text{SSR} = \sum_i \frac{T_i^2}{n_i} = \frac{(GT)^2}{N}$$

$$= \left[\frac{(58)^2}{4} + \frac{(37)^2}{4} + \frac{(50)^2}{4}\right] - \text{C.F}$$

$$= (841 + 342.25 + 625) - 1752.08$$

$$= 1808.25 - 1752.08$$

$\therefore$ 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	$\frac{56.17}{2} = 28.08$	$F_{(2,6)} = \frac{28.09}{20.75} = 1.353$
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.

5. 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.

Block	Soya bean				Total
	Type A	Type B	Type C	Type D	
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 $= \frac{58.25}{3} = 19.41$

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	

$$\text{F-Ratio} = \frac{\text{Mean square between soya bean type}}{\text{Mean square within blocks}}$$

$$= \frac{19.42}{2.08} = 9.33$$

[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.

6. 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				Total
	A	B	C	D	
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)

$$\text{SSR} = \sum_j \frac{T_i^2}{n_i} - \frac{(GT)^2}{N} = \frac{(9)^2}{4} + \frac{(-1)^2}{4} + \frac{(-8)^2}{4} - 0$$

$$= 20.25 + 0.25 + 16 - 0 = 36.5$$

$\therefore$ 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	$\frac{36.5}{2} = 18.25 = (S_m^2)$	$\frac{18.25}{23.25} = 0.784$
Residual or error	139.5	$(C-1)\times(r-1)$ $3 \times 2 = 6$	$\frac{139.5}{6} = 23.25 - (S_e^2)$	
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 :

Definition

If a set of events $A_1, A_2, ..., A_n$ are observed to occur with frequencies $Q_1, O_2, ..., O_n$ respectively and according to probability rules $A_1, A_2, ..., A_n$ are expected to occur with frequencies $E_1, E_2, ...E_n$ respectively with $Q_1, O_2, ..., O_n$ are called observed frequencies and $E_1, E_2, ...E_n$ are called expected frequencies.

If O_i (i = 1, 2,.....,n) is a set of observed (experimental) frequencies and E_i(i = 1, 2,.....,N) is the corresponding set of expected (theoretical) frequencies, then

χ^2 is defined $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$ with (n – 1) degrees of freedom.

χ^2 is used to test whether differences between observed and expected frequencies are significant.

Note :

If the data is given in a series of numbers then degrees of / freedom = n – 1.

In case of Binomial distribution, d.f. = n – 1.

In case of Poison distribution, d.f. = n – 2

In case of Normal distribution, d.f. = n – 3

Chi-square Distribution is an important continuous probability distribution and it is used in both large and small tests. In chi-square tests, χ^2 – distribution is mainly used

(i) to test the goodness of fit,

(ii) to test the independence of attributes,

(iii) to test if the population has a specified value of the variance σ^2.

Q5. Explain the Assumptions and Uses of Chi-square distribution.

Ans :

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.

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

$\chi^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sigma_0^2} = \dfrac{ns^2}{\sigma_0^2} \sim$ a Chi Square Distribution with (n–1) D.F. The value of $s^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ 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.

4.2.2 Test for Goodness of Fit

Q7. Explain briefly about Test for Goodness of fit.

Ans :

We use this test to decide whether the discrepancy between theory and experiment is significant (or) not i.e., to test whether the difference between the theoretical and observed values can be attributed to chance or not.

Let the Null Hypothesis H_0 be that there is no significant difference between the observed values and the corresponding expected values.

The Alternative Hypothesis H_1 is that the above difference is significant.

Let $O_1, O_2, \ldots, O$ be a set of observed frequencies and $E_1, E_2, \ldots, E_n$ the corresponding set of expected frequencies. Then the test statistic χ^2 is given by

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right] = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + ... + \frac{(O_n - E_n)^2}{E_n}$$

Assuming that H_0 is true, the test statistic χ^2 follows Chi-square distribution with (n – l)d.f., where

$$\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i \text{ (or) } \sum_{i=1}^{n}(O_i - E_i) = 0$$

Conclusion

If the calculated value of χ^2 > tabulated value of χ^2 at α level, the Null Hypothesis H_0 is rejected. Otherwise, H_0 is accepted.

χ^2 test enables us to ascertain how well the theoretical distributions such as Binomial, Poisson, Normal, etc., fit the distributions obtained from sample data.

Conditions of Validity

Following are the conditions which should be satisfied before χ^2 test can be applied.

1. The sample observations should be independent.
2. N, the total frequency is large, i.e, > 50.
3. The constraints on the cell frequencies, if any, are linear.
4. No theoretical (or expected) frequency should be less than 10. If small theoretical frequencies occur, the difficulty is overcome by grouping 2 (or) more classes together before calculating (O – E). Note that the degrees of freedom is determined with the number of classes after regrouping.

PROBLEMS

7. **A die is thrown 264 times with the following results. Show that the die is used. Given $\chi^2_{0.05}$ = 11.07 for 5 d.f]**

No.appeared on the die	**1**	**2**	**3**	**4**	**5**	**6**
Frequenc	**40**	**32**	**28**	**58**	**54**	**52**

Sol :

Null hypothesis H_0 : The die is unbiased.

Alternate hypothesis H_i: The die is based.

The expected frequency of each of the numbers 1, 2, 3, 4, 5, 6 is

Observed frequency(O_i)	**Expected frequency(E_i)**	**(O_i – E_i)**	**(O_i – E_i)^2**	$\frac{(O_i - E_i)^2}{E_i}$
40	44	4	16	0.3636
32	44	12	144	3.2727
28	44	16	256	5.8181
58	44	14	196	4.4545
54	44	10	100	2.2727
52	44	8	64	1.4545
264	264			17.6361

$$\therefore \quad \chi^2 = \Sigma \frac{(O_i - E_i)}{E_i} = 17.6362$$

The number of degrees of freedom = n – 1 = 5

The tabulated value of %^{2} for 5 d.f at 5% level = 11.07

Since calculated χ^2 > tabulated χ^2 > we reject the null hypothesis $H_{()}$.

i.e., we reject the hypothesis that the die is unbiased. Hence the die is biased.

8. 200 digits were chosen at random from a set of tables. The frequencies of the digits are shown below :

Digit	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Frequency	**18**	**19**	**23**	**21**	**16**	**25**	**22**	**20**	**21**	**15**

Use the chi-square test to assess the correctness of the hypothesis that the digits were distributed in equal number in the tables from which these were chosen.

Sol :

Null Hypothesis H_0 : The digits were distributed in equal number in the tables.

Alternative Hypothesis H_i : The digits were distributed not in equal numbers in the tables.

The expected frequency of each digit $= \frac{200}{10} = 20$

Calculation for χ^2

Observed frequency(O_i)	Expected frequency(E_i)	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
18	20	2	4	0.2
19	20	1	1	0.05
23	20	3	9	0.45
21	20	1	1	0.05
16	20	4	16	0.8
25	20	5	25	1.25
22	20	2	4	0.2
20	20	0	0	0
21	20	1	1	0.05

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 4.3$$

$\therefore$ Calculated $\chi^2 = 4.3$

Tabulated χ^2 for 10 – 1 = 9 d.f. at 5% level of significance is 16.919. Since calculated χ^2 < tabulated χ^2, we accept the null hypothesis H_0. i.e., the digits are distributed in equal number in the tables.

4.2.3 Test for Independence of Attributes

Q8. Explain briefly about Test for Independence of Attributes.

Ans :

An attribute means a quality or characteristic. Example of attributes are drinking, smoking, blindness, honesty, beauty etc.

An attribute may be marked by its presence (position) or absence in a number of a given population. Let the observations be classified according to two attribute and the frequencies O_i in the different categories be shown in a two-way table called contingency table. We have to test on the basis of cell frequencies whether the two attributes are independent or not. We take the Null - Hypothesis H_0 that there is no association between the attributes i.e., we assume that the two attributes are independent. The expected frequencies (E_i) of any cell $= \dfrac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$

The test statistic $\chi^2 = \sum_i \left[\dfrac{(O_i - E_i)^2}{E_i}\right]$ approximately follows Chi-square distribution with d.f. = (No. of rows – 1) × (No. of columns – 1)

If the calculated value of χ^2 is less than the table value at a specified level (generally 5%) of significance, the hypothesis holds good i.e., the attributes are independent and do not bear any association. On the other hand, if the calculated value of is greater than the table value at a specified level of significance, we say that the results of the experiment do not support the hypothesis, in other words, the attributes are associated.

Let us consider two attributes A and B. A is divided into two classes and b is added into two classes. The various cell frequencies can be expressed in the lowing table known as 2 × 2 contigency table.

A	a	b
B	c	d

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	$N = a+b+c+d$

The expected frequencies are given by

$E(a) = \dfrac{(a+c)(a+b)}{N}$	$E(b) = \dfrac{(b+d)(a+b)}{N}$	a + b
$E(c) = \dfrac{(a+c)(c+d)}{N}$	$E(d) = \dfrac{(b+d)(c+d)}{N}$	c + d
a + c	b + d	N = a + b + c + d

The value of χ^2 is given by $\chi^2 = \dfrac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ where

N = a + b + c + d with d.f. = (2 – 1) (2 – 1) = 1. We use this formula when the expected frequencies are in fractions (or decimals).

PROBLEMS

9. The following table gives the classification of 100 workers according sex and nature of work. Test whether the nature of work is independent of the sex of the worker.

	Stable	Unstable	Total
Male	40	20	60
Females	10	30	40
Total	50	50	100

Sol : **(Aug.-21)**

Null Hypothesis H_0 : The nature of work is independent of the sex of the workers.

Alternate Hypothesis H_i : The nature of work is dependent of the sex of the workers.

Expected frequencies are given in the table.

$\frac{50 \times 60}{100} = 30$	$\frac{50 \times 60}{100} = 30$	60
$\frac{50 \times 40}{100} = 20$	$\frac{50 \times 40}{100} = 20$	40
50	50	100

Calculation of χ^2 :

O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
40	30	10	100	03.333
20	30	10	100	3.333
10	20	10	100	5.000
30	20	10	100	5.000
100	100	40	$\Sigma \frac{(O_i - E_i)^2}{E_i}$	16.66

$$\therefore \quad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 16.66$$

$\therefore$ Calculated $\chi^2 = 16.66$

Tabulated value of χ^2 for (2 – 1) (2 – 1) = 1 d.f. at 5% level of significance is 3.84.

since calculated $\chi^2 >$ tabulated χ^2, we reject the null hypothesis H_0, i.e., the nature of work is not independent of the sex of the workers.

i.e., there is differences in the nature of work on the basis of sex.

10. 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				Total
	Below Average	Average	Above Average	Genius	
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

Sol : **(Aug.-21)**

(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	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
86	84	2	4	0.0476
60	62	2	4	0.0645
44	46	2	4	0.0869
10	8	2	4	0.5
40	42	2	4	0.0952
33	31	2	4	0.129
25	23	2	4	0.1739
2	4	2	4	1
				2.0971

(v) New $\chi^2_{Cal} = \Sigma \dfrac{(O_i - E_i)^2}{E_i} = 2.0971$

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.

4.3 Correlation

Q9. Define Correlation. Explain the significance of Correlation.

Ans :

Meaning

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.

Definitions

(i) **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".

(ii) **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

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, rainfall 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 correlations?

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.

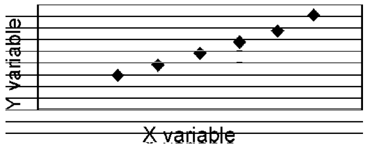


Fig.: Perfect Positive Correlation (r = +1)

(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.

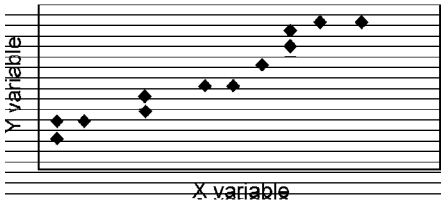


Fig.: Very High Positive Correlation (r = nearly + 1)

(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.

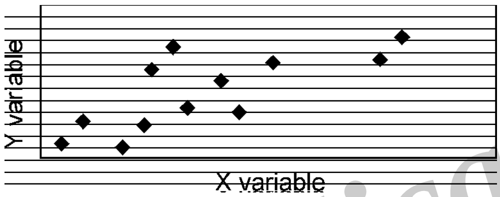


Fig.: Very Low Positive Correlation (r = near to +0)

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.

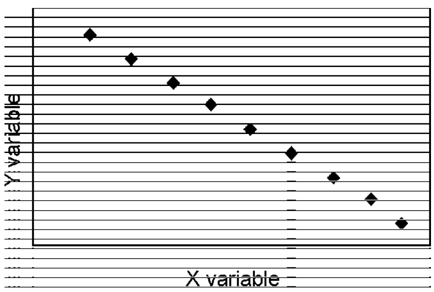


Fig.: Perfect Negative Correlation (r = –1)

(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.

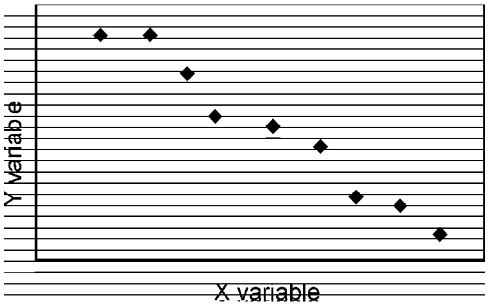


Fig.: Very High Negative Correlation (r = near to –1)

(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.

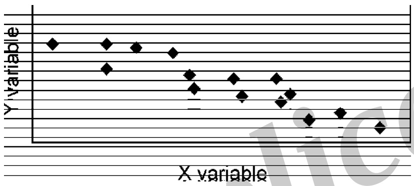


Fig.: Very Low Negative Correlation

(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.

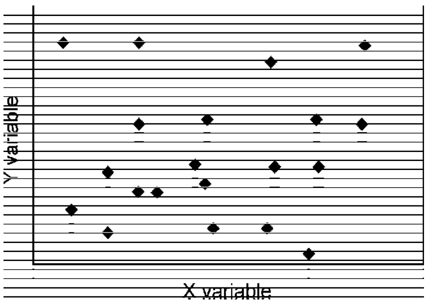


Fig.: No Correlation (r = 0)

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.

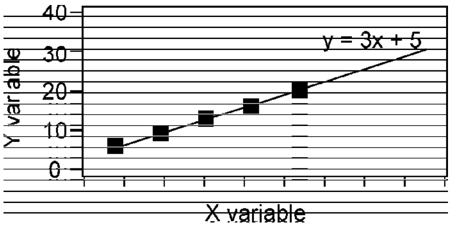


Fig.: Linear Correlation

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

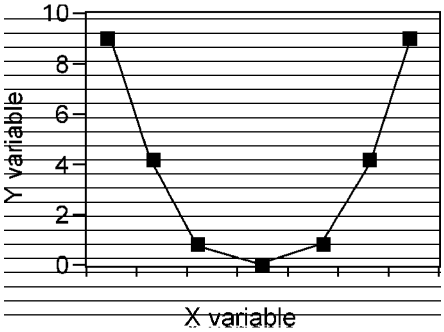


Fig.: Non Linear Correlation

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.

S.No.	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 above but 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

4.4 Karl Pearson's Coefficient of Correlation

Q13. What is Karl Pearson's Coefficient of Correlation? Explain properties of Coefficient of Correlation.

Ans : **(Imp.)**

Meaning

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

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.

Q14. Explain merits and demerits of Coefficient of Correlation.

Ans :

Merits

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

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.

Q15. 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 Method

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 ± 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 ± P.E. i.e. 0.8 ± 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

11. Find if there is any significant correlation between the heights and weights given below:

Height in inches	57	59	62	63	64	65	55	58	57
Weight in lbs	113	117	126	126	130	129	111	116	112

Sol :

Computation of coefficient of correlation

Height in inches x	Deviation Mean(60) $X = x - \bar{x}$	Square of deviations X^2	Weight in lbs y	Deviations from Mean $Y = y - \bar{y}$	Square of deviations deviations	Product of deviations X and Y series(XY)
57	–3	9	113	–7	49	21
59	–1	1	117	–3	9	3
62	2	4	126	6	36	12
63	3	9	126	6	36	18
64	4	16	130	10	100	40
65	5	25	129	9	84	45
55	–5	25	111	–9	81	45
58	–2	4	116	–4	16	8
57	–3	9	112	–8	64	24
540	0	102	1080	0	475	216

$$\bar{X} = \frac{540}{9} = 60 \qquad \bar{Y} = \frac{1080}{9} = 120$$

$$\text{Coefficient of correlation } r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \times \Sigma Y^2}}$$

$$\therefore \quad r = \frac{216}{\sqrt{102 \times 475}} = 0.98$$

12. Psychological tests of intelligence and of engineering ability were applied to 10 students. Here is a record of ungrouped data showing intelligence ratio (I.R) and Engineering ratio (E.R) calculate the co-efficient of correlation.

Student	A	B	C	D	E	F	G	H	I	J
I.R	105	104	102	101	100	99	98	96	93	92
E.R	101	103	100	98	95	96	104	92	97	94

Sol :

We construct the following table :

Student	X	Intelligence Ratio $x - \bar{x} = X$	Y	Engineering Ratio $y - \bar{y} = Y$	X^2	Y^2	XY
A	105	6	101	3	36	9	18
B	104	5	103	5	25	25	25
C	102	3	100	2	9	4	6
D	101	2	98	0	4	0	0
E	100	1	95	–3	1	9	–3
F	99	0	96	–2	0	4	0
G	98	–1	104	6	4	36	–6
H	96	–3	92	–6	1	36	18
I	93	–6	97	–1	36	1	6
J	92	–7	94	–4	49	16	28
Totals	990	0	980	0	170	140	92

From this table, we have mean of x, i.e., $\bar{x} = 990/10 = 99$;

mean of y, i.e., $\bar{y} = 980/10 = 98$, $\Sigma X^2 = 170$, $\Sigma Y^2 = 140$ and $\Sigma XY = 92$

Substituting these values in $r = \dfrac{\Sigma XY}{\sqrt{(\Sigma X^2 \times \Sigma Y^2)}} = \dfrac{92}{\sqrt{(170 \times 140)}} = \dfrac{92}{154.3} = 0.59$

13. Find a suitable coefficient of correlation for the following data :

Fertiliser used (tonnes)	15	18	20	24	30	35	40	50
Productivity (tonnes)	85	93	95	105	120	130	150	160

Sol :

Computation of coefficient of correlation

Fertiliser used x	Deviation from assumed mean $X = x - \bar{X}$ $= x - 29$	Square of deviation X^2	Productivity y	Deviation from assumed mean $Y = y - \bar{Y}$ $= y - 119$	Square of deviation Y^2	Productivity of deviation XY
15	–14	196	85	–34	1156	476
18	–11	121	93	–26	676	286
20	–9	81	95	–24	576	216
24	–5	25	105	–14	196	70
30	1	1	120	1	1	1
35	6	36	130	11	121	66
40	11	121	150	31	961	341
50	21	441	160	41	1681	861
	$\Sigma X = 0$	$\Sigma X^2 = 1022$		$\Sigma Y = -14$	$\Sigma Y^2 = 5368$	$\Sigma XY = 2317$

Here N = 8,

$$\therefore \quad r = \frac{\Sigma XY \times N - (\Sigma X)(\Sigma Y)}{\sqrt{(\Sigma X)^2 \times N - (\Sigma X)^2 \times (\Sigma Y)^2 \times N - (\Sigma Y)^2}}$$

$$= \frac{2317}{\sqrt{1022 \times 5343.5}} = \frac{2317}{2336.89} = 0.99$$

4.4.1 Spearman's Rank Correlation

Q16. Explain about rank correlation and its features.

(OR)

What is Spearman's Rank Correlation?

Ans : **(May-22)**

Meaning

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

- 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.

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 $R = 1 - \frac{6\Sigma D^2}{N^3 - N}$

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 $R = 1 - \frac{6\Sigma D^2}{N^3 - N}$

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

$$R = 1 - \frac{6\left\{\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + ...\right\}}{N^3 - N}$$

Q17. 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

14. The ranking of 10 students in two subjects A and B are as follows :

A	B
6	3
5	8
3	4
10	9
2	1
4	6
9	10
7	7
8	5
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
			$\Sigma D^2 = 36$

$$R = 1 - \frac{6\Sigma D^2}{N^3 - N} = 1 - \frac{6 \times 36}{10^3 - 10}$$

$$= 1 - \frac{216}{990} = 0.782$$

15. 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	$(R_x - R_y)^2$ 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
					$\Sigma D^2 = 62$

$$R = 1 - \frac{6\,\Sigma D^2}{N^3 - N} = 1 - \frac{6 \times 62}{7^3 - 7} = 1 - 1.107 = -0.107.$$

16. 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).

$$R = 1 - \frac{6\left\{\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right\}}{N^3 - N}$$

$$R = 1 - \frac{6\left\{134 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^3 - 3)\right\}}{10^3 - 10}$$

$$= 1 - \frac{6[134 + 2 + .5 + .5 + .5 + 2]}{990}$$

$$= 1 - \frac{6(139.5)}{990} = 1 - \frac{837}{990} = 1 - .845 = 0.155$$

17. What is Spearman's Rank Correlation? Solve the following problem on Spearman's Rank Correlation.

Candidate	Judge P	Judge Q
1	**20**	**24**
2	**23**	**24**
3	**25**	**22**
4	**24**	**26**
5	**26**	**30**

Sol : **(May-22)**

Candidate	Judge P	Judge Q	R_1	R_2	D	D^2
1	20	24	1	2.5	–1.5	2.25
2	23	24	2	2.5	–0.5	0.25
3	25	22	4	1	3	9
4	24	26	3	4	–1	1
5	26	30	5	5	0	0
						12.5

$$1-\frac{6\left[\Sigma D^2+\frac{1}{12}(m^3-m)\right]}{N^3-N}$$

$$1-\frac{6\times 12.5+\frac{1}{12}(2^3-2)}{(50)^3-5}$$

$$1-\frac{75+0.5}{125-5}$$

$$1-\frac{75.5}{120}$$

$$1-0.629 = 0.371$$

4.5 Linear and Multiple Regression Analysis

Q18. 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

(i) “Regression is the measure of the average relationship between two or more variables in terms of the original units of data.”

(ii) **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."

(iii) **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

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.

Q19. 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

Q20. 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.

Q21. 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.

Q22. 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,

$$y = ao + ai\ x + a_2 x^2 + a3X^3$$

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

Q23. Explain briefly 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

18. From the following data obtain the two regression equations and calculate the correlation co-efficient.

x	2	4	6	8	10	12	14	16	18
y	18	16	20	24	22	26	28	32	30

Calculate the value of y when x = 6.2

Sol :

(i) X on Y

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$

(ii) Y on X

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$\bar{X} = \frac{\Sigma x}{n} = \frac{90}{9} = 10$$

$$\bar{Y} = \frac{\Sigma y}{n} = \frac{216}{9} = 24$$

Calculation of Regression Equation and Correlation Coefficient

X	Y	$x - \bar{x}$ x	$y - \bar{y}$ y	x^2	y^2	xy
2	18	–8	–6	64	36	48
4	16	–6	–8	36	64	48
6	20	–4	–4	16	16	16
8	24	–2	0	4	0	0
10	22	0	–2	0	4	0
12	26	2	2	4	4	4
14	28	4	4	16	16	16
16	32	6	8	36	64	48
18	30	8	6	64	36	48
90	216	0	0	240	240	228

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{228}{240} = 0.95$$

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{228}{240} = 0.95$$

Equation X on Y

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$

$$X - 10 = 0.95\,(Y - 24)$$

$$X = 0.95\,Y - 22.8 + 10$$

$$X = 0.95\,Y - 12.8$$

Equation Y on X

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$Y - 24 = 0.95\,(X - 10)$$

$$Y - 24 = 0.95,\ X - 95 + 24$$

$$Y = 0.95\,x - 14.5$$

Regression equation y on x = 6.2

$$y = 0.95x - 14.5$$

$$y = 0.95(6.2) - 14.5$$

$$y = -8.61$$

19. Given :

$\Sigma x=56$, $\Sigma y=40$, $\Sigma x^2=524$, $\Sigma y^2=256$, $\Sigma xy=364$, $N = 8$

(i) Find the two Regression equations and

(ii) The Correlation Coefficient.

Sol :

We have

$$\bar{X} = \frac{\Sigma x}{N} = \frac{56}{8} = 7; \quad \bar{y} = \frac{\Sigma y}{N} = \frac{40}{8} = 5$$

byx = co-efficient of regression of y on x = $\dfrac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N\left(\Sigma y^2\right) - (\Sigma x)^2}$

$$\frac{8(364)-(56)(40)}{8(524)-(56)^2} = \frac{2912-2240}{4192-3136} = \frac{672}{1056}$$

$$\boxed{\text{byx} = 0.6363}$$

bxy = co-efficient of regression of x on y = $\dfrac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N\left(\Sigma y^2\right) - (\Sigma y)^2}$

$$\frac{8(364)-(56)(40)}{8(256)-(40)^2} = \frac{2912-2240}{2048-1600} = \frac{674}{448} = \boxed{bxy = 1.504}$$

(i) Two Regression equations

Regression equation x on y

$$(x-\bar{x}) = b_{xy}(y-7)$$

(x –7) = 1.504 (y – 5)

(x – 7) = 1.504 (y) – 1.504 (5)

x = 1.504 (y) – 7.522 + 7

x = 1.504 (y) – 0.522 .... (1)

Regression equation y on x

$$(y-\bar{y}) = b_{yx}(x-\bar{x})$$

(y – 5) = 0.6363 (x – 7)

y – 5 = 0.6363 (x) – 0.6363(7)

y = 0. 6363 (x) – 4.4541 + 5

y = 0.6363 (x) + 0.5459 ....(2)

(ii) The correlation co-efficient γ_{xy} between x and y is given by

$$\gamma_{xy}^2 = b_{yx}.b_{xy} = (0.6363)(1.504)$$

$$r_{xy}^2 = 0.9569$$

$$\boxed{\gamma_{xy} = 0.9782}$$

20. Following are the marks in Statistics and English in an Annual Examination.

Particular	Statistics (X)	English (Y)
Mean	40	50
Standard Derivation	10	16
Co-efficient Correlation	0.5	

(i) Estimate the score of English, when the score in Statistics is 50.

(ii) Estimate the score of statistics, when the score in English is 30.

Sol :

Given mean of X denoted as $\bar{X}$ = 40.

Given mean of Y denoted as $\bar{Y}$ = 50.

SD of X denoted as σ_x = 10.

SD of Y denoted as σ_y = 16.

Coefficient of correlation denoted as r = 0.5

Regression Equation X on Y

$$[X - \bar{X}] = [r]\left[\frac{\sigma x}{\sigma y}\right][y - \bar{y}]$$

$$X - 40 = [0.5]\left[\frac{10}{16}\right][Y - 50]$$

$$X - 40 = [0.5]\ [0.625]\ [Y - 50]$$

$$X - 40 = [0.3125]\ [Y - 50]$$

$$X - 40 = 0.3125y - 15.625$$

$$X = 0.3125y - 15.625 + 40$$

$$X = 0.3125y + 24.375$$

Regression Equation Y on X

$$[Y - \bar{Y}] = [r]\left[\frac{\sigma y}{\sigma x}\right][X - \bar{X}]$$

$$[Y - 50] = [0.5]\left[\frac{16}{10}\right][X - 40]$$

$$Y - 50 = [0.5]\ [1.6]\ [X - 40]$$

$$Y - 50 = (0.8)\ (X - 40)$$

$$Y - 50 = 0.8X - 32$$

$$Y = 0.8\,X - 32 + 50$$

$$Y = 0.8\,X + 18$$

Estimation of English (Y) when Statistics (X) is 50

$$Y = 0.8\,X + 18$$

$$= 0.8\,(50) + 18$$

$$= 40 + 18$$

$$\therefore Y = 18 \text{ marks.}$$

Estimation of statistics (X) when English (Y) is 30

$$X = 0.3125\,Y + 24.375$$

$$= 0.3125(30) + 24.375$$

$$= 9.375 + 24.375$$

$$X = 33.75 \text{ marks.}$$

Q24. Define multiple regression analysis. State its objectives and assumptions.

Ans : **(May-22)**

Meaning

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

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

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.

Q25. Explain briefly about:

(a) Regression coefficients in multiple regression

(b) R^2 in multiple regression.

Ans : **(May-22)**

(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,

$$R^2 = \frac{SSR}{SST}$$

Q26. Differentiate between Correlation and Regression.

(OR)

What are the differences between Correlation and Regression?

Ans : **(June-18, Imp.)**

S.No.	Nature	Correlation	Regression
1.	**Meaning**	Correlation is a statistical measure which determines co-relationship (or) association association of two variables.	Regression describe how an independent variable is numerically related to the dependent variable.
2.	**Usage**	To represent linear relationship between two variables.	To fit a best line and estimate one variable on the basis of another variable. Both
3.	**Dependent and Independent variables**	No difference	variables are different
4.	**Indicates**	Correlation coefficient indicates the extent to which two variables move together.	Regression indicates the impact of a unit change in the known variable (x) on the estimated variable (Y).
5.	**Objective**	To find a numerical value expressing the relationship between variables. variable.	To estimate values of random variable on the basis of the values of fixed

Q27. Explain clearly the differences between simple linear and multiple linear regressions.

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

S.No.	Simple Regression	Multiple Regression
1.	Dependent variable Y predicted from one independent variable X	Dependent variable Y predicted from a set of independent variables (X_1, X_2 ... X_k)
2.	Regression coefficient variable.	Regression coefficient for each independent
3.	r^2 proportion of variation in dependent variable Y predictable from X.	R^2 proportion of variable in dependent variable Y predictable by set of independent variables (X's)

4.6 Discriminant Analysis

Q28. What is Discriminant Analysis? Explain the objectives of discriminant analysis.

Ans : **(Imp.)**

Meaning

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

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.

Q29. 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.

Q30. 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

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.

Q31. 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.

Q32. Explain the steps for conducting the Discriminant Analysis.

Ans :

Th following are the steps involved in Discriminant Analysis:

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.

Problem Formulation

↓

Coefficient Estimation

↓

Significance Determination

↓

Results Interpretation

↓

Validation

Fig.: Steps for Conducting Discriminant Analysis

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

Q33. Define Exploratory Factor Analysis. What are the objectives of Exploratory Factor Analysis?

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

Meaning

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

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

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

$$X_i = A_{i1} F_1 + A_{i2} F_2 + A_{i3} F_3 + \ldots\ldots A_{im} F_m + V_i U_i$$

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.

Q34. What are the advantages and disadvantages of factor analysis?

Ans : **(Imp.)**

Advantages

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

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.

Q35. 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:

- Centroid method
- Principal components analysis
- Maximum likelihood method
- Unweighted least squares method
- Generalized least squares method
- Principal axis factoring method
- Alpha factoring method
- 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.

Q36. 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.

Q37. 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 :

(i) 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.

(ii) 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.

(iii) 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.

(iv) 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.

(v) 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,

i) Orthogonal rotation

ii) Varimax rotation and

iii) Oblique rotation

(i) 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.

(ii) 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.

(iii) 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.

Q38. 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.

Short Question and Answers

1. What is an ANOVA?

Ans :

Meaning

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 population 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.

2. Chi-Square Distribution.

Ans :

Definition

If a set of events $A_1, A_2, ...,A_n$ are observed to occur with frequencies $Q_1, O_2,.....,O_n$ respectively and according to probability rules $A_1, A_2,.....,A_n$ are expected to occur with frequencies $E_1, E_2,.....E_n$ respectively with $Q_1, O_2,.....,O_n$ are called observed frequencies and $E_1, E_2,.....E_n$ are called expected frequencies.

If O_i (i = 1, 2,.....,n) is a set of observed (experimental) frequencies and E_i(i = 1, 2,.....,N) is the corresponding set of expected (theoretical) frequencies, then

$$\chi^2 \text{ is defined } \chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \text{ with } (n-1)$$

degrees of freedom.

χ^2 is used to test whether differences between observed and expected frequencies are significant.

Note :

If the data is given in a series of numbers then degrees of / freedom = n – 1.

In case of Binomial distribution, d.f. = n – 1.

In case of Poison distribution, d.f. = n – 2

In case of Normal distribution, d.f. = n – 3

Chi-square Distribution is an important continuous probability distribution and it is used in both large and small tests. In chi-square tests, χ^2 – distribution is mainly used

(i) to test the goodness of fit,

(ii) to test the independence of attributes,

(iii) to test if the population has a specified value of the variance σ^2.

3. Use of Chi-Square Test.

Ans :

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).

4. Define Correlation.

Ans :

Meaning

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.

Definitions

(i) **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".

(ii) **According to Tippet,** "The effects of correlation are to reduce the range of uncertainty of our prediction".

5. Significance of Correlation.

Ans :

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, rainfall 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.

6. Demerits.

Ans :

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. Spearman's Rank Correlation.

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 :

$$R = 1 - \frac{6\Sigma D^2}{N(N^2-1)} \text{ or } 1 - \frac{6\Sigma D^2}{N^3 - N}$$

8. 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."

(i) **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."

(ii) **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."

9. 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.

10. Define 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

Exercise Problems

1. Find out the coefficient of correlation of Karl Pearson from the following data:

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

[Ans: r = 0.603]

2. Calculate the coefficient of correlation and find its probable error from the following data.

X	7	6	5	4	3	2	4
Y	18	16	14	12	10	6	8

[Ans. r = 0.9643, P.E. 0.0179]

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]

Choose the Correct Answer

1. Analysis of variance is a statistical method of comparing the of several populations. [a]
   (a) Means (b) Variances
   (c) Standard Deviations (d) None of The Above
2. The ______ sum of squares measures the variability of the observed values around their respective treatment means. [a]
   (a) Error (b) Total
   (c) Treatment (d) Interaction
3. Which of the following is an assumption of one-way ANOVA comparing samples from three or more experimental treatments? [d]
   (a) The samples associated with each population are randomly selected and are independent from all other samples
   (b) The response variable within each of the k populations have equal variances
   (c) All the response variables within the k populations follow a normal distributions
   (d) All of the above
4. When the k population means are truly different from each other, it is likely that the average error deviation: [a]
   (a) is relatively small compared to the average treatment deviations
   (b) is about equal to the average treatment deviation
   (c) is relatively large compared to the average treatment deviations
   (d) none of the above
5. In a study, subjects are randomly assigned to one of three groups: control, experimental A, or experimental B. After treatment, the mean scores for the three groups are compared. The appropriate statistical test for comparing these means is: [a]
   (a) The Analysis of Variance (b) The Correlation Coefficient
   (c) Chi Square (d) The T-Test
6. When conducting an ANOVA, FDATAwill always fall within what range? [a]
   (a) between 0 and infinity (b) between 0 and 1
   (c) between negative infinity and infinity (d) between 1 and infinity
7. Which of the following are types of correlation? [d]
   (a) Positive and Negative (b) Simple, Partial and Multiple
   (c) Linear and Nonlinear (d) All of the above
8. Which of the following is true for the coefficient of correlation? [c]
   (a) The coefficient of correlation is not dependent on the change of scale
   (b) The coefficient of correlation is not dependent on the change of origin
   (c) The coefficient of correlation is not dependent on both the change of scale and change of origin
   (d) None of the above
9. Which of the following statements is true for correlation analysis? [c]
   (a) It is a bivariate analysis (b) It is a multivariate analysis
   (c) It is a univariate analysis (d) Both a and c
10. If the values of two variables move in the same direction, ________. [d]
    (a) The correlation is said to be non-linear (b) The correlation is said to be linear
    (c) The correlation is said to be negative (d) The correlation is said to be positive

Fill in the Blanks

1. _______ is used in education, industry, business, psychology field mainly in their experiment design.
2. _______ Distribution is an important continuous probability distribution and it is used in both large and small tests.
3. An _______ means a quality or characteristic.
4. _______ analysis is the statistical tool we can use to describe the degree to which one variable is linearly related to another.
5. The value of the coefficient of correlation should lie between _______ and _______.
6. The _______ of the coefficient of correlation helps in interpretation.
7. The term 'regression' was first used by _______ in 1877
8. _______ helps in establishing a functional relationship between two or more variables.
9. _______ regression analysis represents a logical extension of two-variable regression analysis.
10. Factor Analysis was first used by _______.

ANSWERS

1. Anova
2. Chi-square
3. Attribute
4. Correlation
5. +1, –1
6. probable error
7. Sir Francis Galton
8. Regression analysis
9. Multiple
10. Charles spearman

One Mark Answers

1. Two way ANOVA.

Ans :

Two way classification/two factor ANOVA is defined where two independent factors have an effect on the response variable of interest.

2. Define Correlation.

Ans :

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.

3. Positive Correlation.

Ans :

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.

4. Negative Correlation.

Ans :

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.

5. Simple Regression.

Ans :

In statistics, simple regression is the least squares estimator of a linear regression model with a single predictor variable.

6. Discriminant Analysis.

Ans :

A discriminant analysis enables the researcher to classify persons or objects into two or more categories.

UNIT V	**Time Series Analysis and Report Writing:** Components, Models of Time Series, Additive, Multiplicative and Mixed Models, Trend Analysis: Free hand Curve, Semi Averages, Moving Averages, Least Square Methods. **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. 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.

(OR)

What is time series ?

Ans : **(Imp.)**

Meaning

A time series is a statistical data that are collected, observed or recorded at regular intervals of time. The term time series applies, for example, to the data recorded periodically showing the total annual sales of retail stores, the total quarterly value of construction contracts awarded, the total amount of unfilled orders in durable goods industries at the end of each month, weekly earnings of workers in an industrial town, hourly temperature in a particular city.

Definitions

Some of the important definitions of time series, given by different experts are as under:

(i) According to Morris Hamburg, "A time series is a set of statistical observations arranged in chronological order."

(ii) According to Patterson, "A time series consists of statistical data which are collected, recorded observed over successive increments."

(iii) According to Ya-Lun-Chou, "A time series may be defined as a collection of magnitudes belonging to different time periods, of some variable or composite of variables, such as production of steel, per capita income, gross national product, price of tobacco, or index of industrial production."

(iv) According to Wessel and Wellet, "When quantitative data are arranged in the order of their occurrence, the resulting statistical series is called a time series."

(v) According to Spiegel, "A time series is a set of observations taken at specified times, usually at 'equal intervals'. Mathematically, a time series is defined by the values Y_1 Y_2 ... of a variable Y (temperature, closing price of a share, etc.) at times t_1 , t_2 .... Thus Y is a function of t_1 symbolized by Y= F(t)."

(vi) According to Cecil H. Mayers, "A time series may be defined as a sequence of repeated measurement of a variables made periodically through time".

It is clear from the above definitions that time series consist of data arranged chronologically. Thus if we record the data relating to population, per capita income, prices, production, etc., for the last 5, 10, 15, 20 years or some other time period, the series so emerging would be called time series.

It should be noted that the term 'time series' is usually used with reference to economic data and the economists are largely responsible for the development of the techniques of time series analysis. However, the term 'time series' can apply to all other phenomena that are related to time such as the number of accidents occurring in a day, the variation in the temperature of a patient during a certain period, number of marriages taking place during a certain period, etc.

Q2. What are the characteristics of time series ?

Ans : **(Imp.)**

The essentials of Time Series are

i) It must consist of a set of values that are homogeneous. For example, production data for a year and sales data for the next year will not be a time series.

ii) The values must be with reference to time. In other words, in a time series, we have at least 2 variables, with one variable necessarily being time.

iii) The data must be available for a reasonably long period of time.

iv) The gaps between various time values should as far as possible be equal.

v) The values of the second variable should be related to time. For example, the number of people being hired by the BPO industry can be tracked in relation to time. However, if we are talking about average height of students in a class, this data may not have a significant relationship with time and may not constitute a time series.

Q3. What are the objectives of time series ?

Ans :

While it is true that past performance does not necessarily guarantee future results, the quality of forecasts that management can make is strongly related to the information that can be extracted and used from past data. Thus, the objective of time series analysis is to interpret the changes in a given variable with reference to the given situation and attempt to anticipate the future course of events. Analysis of Time Series is done with the following objectives:

i) To evaluate past performance in respect of a particular variable.

ii) To make future forecasts in respect of the particular variable.

iii) To chart short term and long term strategies of the business in respect of the particular variable.

5.1.1 Components

Q4. What are the Components of Time Series ?

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

The changes in the values of a variable related to time can be the result of a large Variety of factors such as change in the tastes and habits of people, change in population, change in cost of production, change in income of people, change in relationship between countries, addition or elimination of competitors, change in climatic conditions, change in Government policy etc. The value of a variable changes due to interaction of such factors. The various factors affecting the values of the given variable may be broadly classified into four categories, commonly known as Components of Time Series.

Each of these components explains some distinct characteristics of change, long term or short term, regular or irregular or non-repetitive. The various components of Time Series can be broadly classified as under:

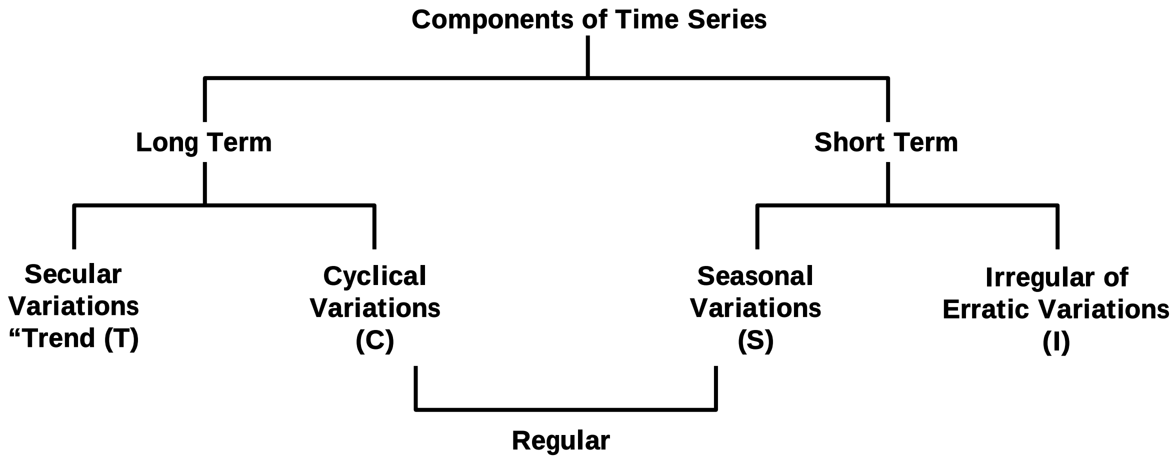


Fig.: Components of Time Series

Thus, the components of Time Series are:

i) Secular Trend or Long Term Movement

ii) Seasonal Variations

iii) Cyclical Variations and

iv) Irregular, Random or Erratic Variations

(i) Secular Trend (or) Long Term Movement

Secular Trend is the basic tendency of a series to grow, decline, remain constant or fluctuate, over a long period of time. The concept of trend does not include short term changes but is concerned with steady movement over a long period. The long term trend movement is the result of forces that experience change very gradually and continuously over a long period of time. They operate in an evolutionary manner and do not reflect sudden changes. For example, the number of people travelling by Airways has gradually and continuously gone up. The number of infant deaths per thousand children bom is a steadily declining trend. This gradual and continuous movement of trend can be attributed to various factors such as increased investment in infrastructure, opening up of the economy through economic reforms, advances in technology, change in demographic profile etc.

The long term trend helps us in determining the direction of change. The growth factor can be estimated with a fair degree of accuracy. It provides indication of what is ahead for a given series. Elimination of trend from the original data helps in understanding the other elements that influence data in the short run.

(ii) Seasonal Variations

Seasonal variations involve patterns of change within a year that tend to get repeated year after year. The trend is fluctuating and repetitive as it peaks and bottoms out at about the same time of the year, every year. Seasonal variations can be detected only if the data is recorded in smaller units of time such as weekly, monthly or quarterly. There are no seasonal variations in a time series where only annual figures are available. These are the result of such factors that uniformly rise and fall in magnitude. For example, the sale of umbrellas peaks in the months of June and July every year. This is on account of the onset of monsoons. The prices of agricultural commodities fall at the time of harvest. The passenger traffic increases substantially in the summer vacations. In all these examples, the movement of the trend is for a short period of time (season). The same movements are repeated in the coming years. Hence, it is easy to forecast the future.

Characteristics: Based on the above discussion, we can list the following characteristics of seasonal variations.

a) They repeat themselves periodically in less than one year's time.

b) These are results of factors that uniformly and regularly rise and fall in magnitude.

c) These variations are periodic and regular. They can be predicted without much difficulty.

Causes: Seasonal Variations can be on account of natural forces or due to manmade conventions. For example, the trend in umbrellas can be attributed to nature, while increase in passenger traffic in summer is manmade. However in either situation, it is very important to study the seasonal variations. A proper understanding of seasonal variations leads to better planning of future operations. If a seasonal upswing in sales is misinterpreted as a genuine increase in demand for the product, the entire operations of the business will get adversely affected.

(iii) Cyclical Variations

The cyclical variations in a series are the recurrent variations whose duration is more than one year. Cyclical variations are regular but not uniformly periodic. It is relatively more difficult to predict the future direction of the series. One complete period that normally lasts from seven to nine years, is termed as a cycle. The most common example of Cyclical Variations is the business cycle. The business cycle passes through four phases of Boom, Recession, Depression and Recovery and it may take more than 10 years to complete the cycle.

Characteristics: The characteristics of a cyclical variation are

(a) There are Oscillating movements above and below the secular tread line.

(b) The fluctuations occur over a period greater than one year.

(c) There is no uniformity in the period of recurrence of movement pattern. One cycle may get completed in 3 years while the next cycle for the same date may takes 8 years.

(d) Cyclical fluctuations are more difficult to measure.

Utility

(a) Knowledge of cyclical variations helps users to understand that they will not be perennially in the same state of boom or depression. They realize the need or different strategies at different times.

(b) Study of cyclical variations helps businesses to predict the turning points ahead of time.

(c) Understanding of cyclical variations is helpful in formulation of policies such as diversification aimed at stabilizing business fluctuations.

Challenges

(a) There is no uniformity in the period of recurrence of movement pattern.

(b) They ,are mixed with Irregular Variations and it is difficult to separate the two.

(iv) Irregular, Random (or) Erratic Variations

In many situations, the value of a variable may be completely unpredictable, changing in a random manner. These fluctuations are the result of such unforeseen and unpredictable forces that operate in absolutely erratic and irregular manner. There is no definite pattern and there is no regular period of time of their occurrence.

They are normally short-term variations caused by non-recurring factors such as floods, famines, revolution (such as green revolution) etc. However, in rare instances, the effects of these random variations are such that they may lead to new cyclical or seasonal movements. Irregular variations are also known as 'episodic' variations and include all variations other than those accounted for by trend, seasonal and cyclical variations.

5.2 Models of Time Series

5.2.1 Additive, Multiplicative and Mixed Models

Q5. 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

Q6. Define Trend Analysis. Explain the purpose of measuring trend.

Ans : **(Imp.)**

Meaning

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.

Purpose

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.

Q7. Explain the uses of trend analysis.

Ans :

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,.

Q8. 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

Q9. 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.

Merits

(i) It is very simple.

(ii) It does not involve any calculations.

(iii) It is very flexible and can be used irrespective of whether the trend is linear or curvi-linear.

(iv) If used by experienced statisticians, it is a better tool to study trend movement compared to other methods using rigid mathematical formulae.

Limitations

(i) It is very subjective. Different persons may draw different lines and reach different conclusions from the same data. Hence, it is not a good forecasting tool.

(ii) If properly attempted, it is very time-consuming effort.

(iii) It requires high levels of experience and expertise to effectively use this method.

Example

Fit a trend line to the following data by the freehand method,

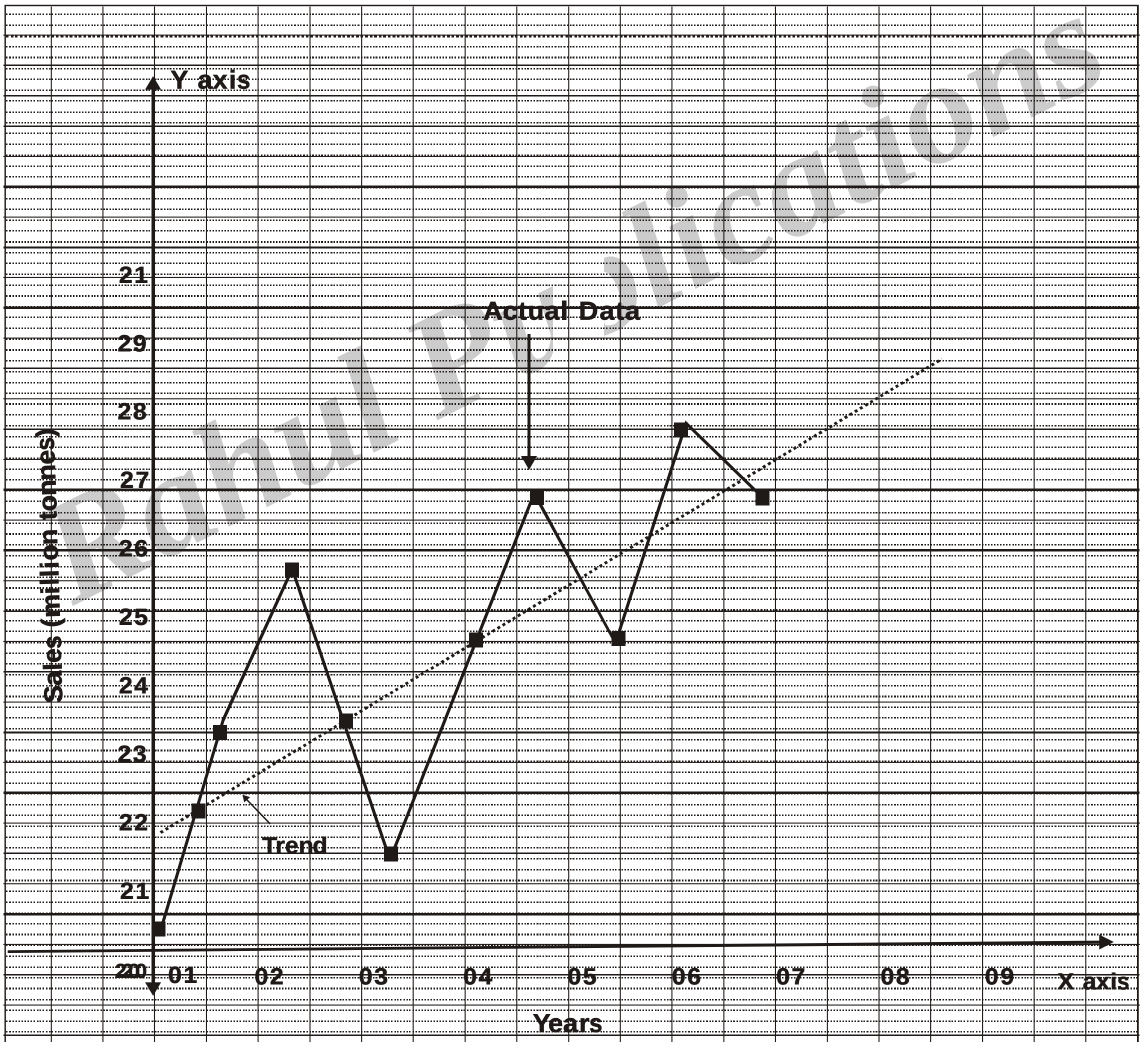
Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Sales (milllion tonnes)	19	22	24	20	23	25	23	26	25

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

Q10. Explain in detail semi average method with an example.

Ans : **(Imp.)**

The following procedure is followed for semi-average method,

- 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.
- Compute the arithmetic mean of time series values for each half separately. These means are called semi-averages.
- Semi averages are plotted as points against the middle point of the respective time period covered by each part.
- The line joining these points gives the straight line trend fitting the given data.

Merits

The following are some of the merits of semi-average method,

- Objectivity
- Ease of apply and understandability
- Extend both ways the line i.e., we can get past and future estimates.

Demerits

Some of the demerits of semi-average method are,

- Linear trend assumption may not exist.
- A men may be questioned.
- 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

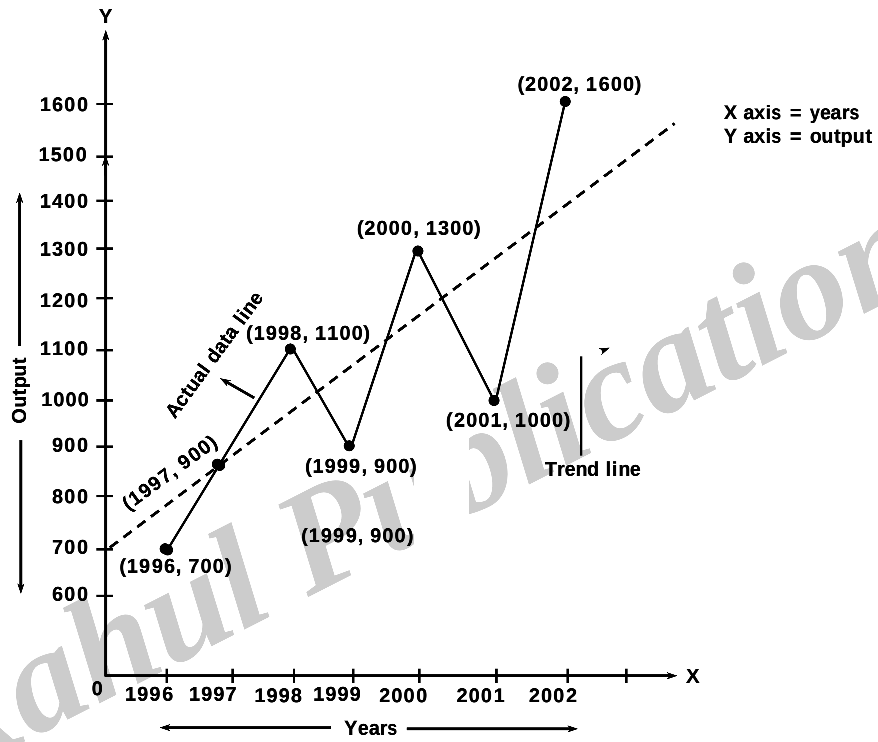
The next step is to plot the semi-averages against the mid-point (middle year) of each time period. Thus, it would , be year 1997 and 2001 respectively.

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

Q11. Discuss the method of moving averages in measuring trend. What are it merits and limitations of moving average method?

Ans : **(Imp.)**

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

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

The merits of moving average are as follows,

(a) Of all the mathematical methods of fitting a trend, this method is the simplest.«

(b) The method is flexible so that even if a few more observations are to be added, the entire calculations are not changed.

(c) If the period of the moving average happens to coincide with the period of the cycle, the cyclical fluctuations are automatically eliminated.

(d) 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

The following are the limitations of moving averages,

(a) 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.

(b) It is difficult to decide the period of moving average since there is no hard and fast rule for the purpose.

(c) Moving average cannot be used in forecasting as it is not represented by any mathematical function.

(d) When the trend is not linear, the moving average lies either above or below the true sweep of the data.

5.3.4 Least Square Methods

Q12. Define least square method. Explain merits and demerits of least square method.

(OR)

What is least square method and explain its advantages and disadvantages?

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 $Y = a + bX$

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

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

Some of the demerits of least squares are as follows,

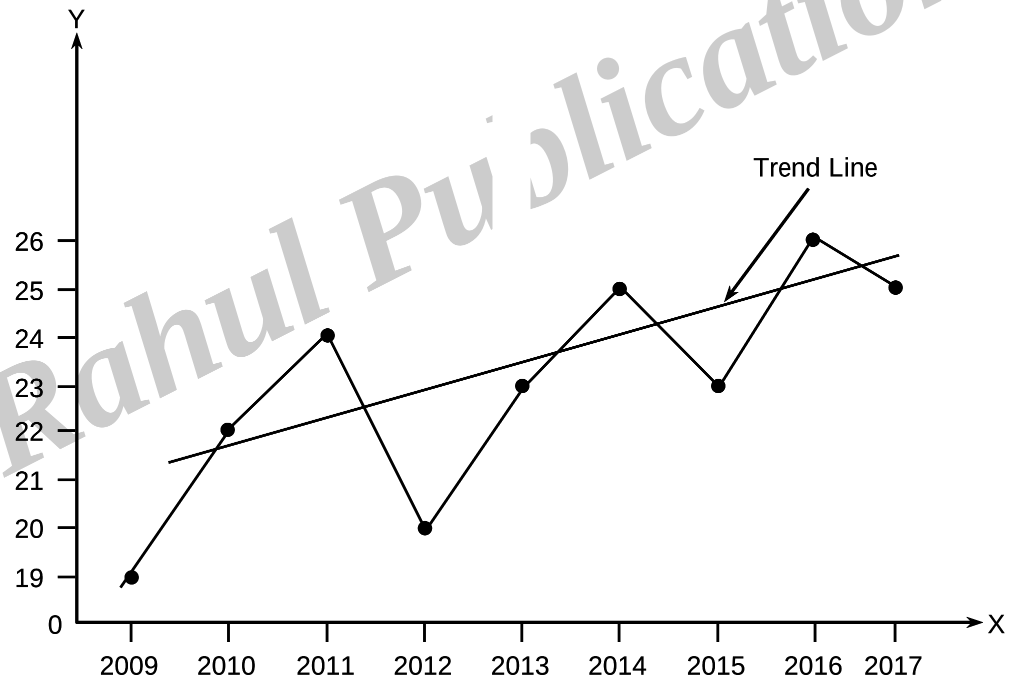
1. Fresh calculations become necessary even if a single new observation is added.
2. Calculations required in this method are quite tedious and time consuming as compared with other methods.
3. Future predictions based on this method completely ignore the cyclical, seasonal and erratic fluctuations.
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

1. Fit a trend line to the following data by the freehand method.

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017
Sales (Rs.)	19	22	24	20	23	25	23	26	25

Sol :



2. From the following data fit a trend line by the method of Semi-Average.

Year :	2012	2013	2014	2015	2016	2017
Output :	20	16	24	30	28	32

Sol :

The trend value of first 3 years calculated as follows

$$\frac{20+16+24}{3} = \frac{60}{3} = 20.$$

The trend value of last 3 years are calculated as follows

$$\frac{30+28+32}{3} = \frac{90}{3} = 30$$

Therefore, the Semi Average are 20 and 30.

3. The following table gives the annual sales (in Rs.'000) of a commodity :

Year	Sales
1990	**710**
1991	**705**
1992	**680**
1993	**687**
1994	**757**
1995	**629**
1996	**644**
1997	**783**
1998	**781**
1999	**805**
2000	**805**

Determine the trend by calculating the 5-yearly moving average.

Sol :

Calculation of Trend by 5-Yearly Moving Average

Year	Sales	5-Yearly Moving Total	5-Yearly Moving Average
1990	710		
1991	705		
1992	680	710 + 705 + 680 + 687 + 757 = 3539	$\frac{3539}{5} = 707.8$
1993	687	705 + 680 + 687+ 757 + 629 = 3458	$\frac{3458}{5} = 691.6$
1994	757	680 + 687 + 757 + 629 + 644 = 3397	$\frac{3397}{5} = 679.4$
1995	629	687 + 757 + 629 + 644 + 783 = 3500	$\frac{3500}{5} = 700$

1996	644	757 + 629 + 644 + 783 + 781 = 3594	$\frac{3594}{5} = 718.8$
1997	783	629 + 644 + 783 + 781 + 805 = 3642	$\frac{3642}{5} = 728.4$
1998	781	644 + 783 + 781 + 805 + 872 = 3885	$\frac{3885}{5} = 777$
1999	805		
2000	872		

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

Week	**Units**	**3 - Yearly Total**	**3 - Yearly Moving 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)

$$S_t = S_{t-1} + \alpha (D_t - S_{t-1})$$

$$S_7 = 17 + 0.2 \ (18 - 17)$$

$$17 + 0.2 \ (1) = 17.2$$

5. Calculate three year moving average for the following data:

Year :	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
Value :	242	250	252	249	253	255	251	257	260	265	262

Sol :

Calculation of Trend by 3 years moving average

Year	Value	3 year moving total	3 year moving average
1950	242	–	–
1951	250	242 + 250 + 252 = 744	744/3 = 248.00
1952	252	250 + 252 + 249 = 751	751/3 = 250.33
1953	249	252 + 249 + 253 = 754	754/3 = 251.33
1954	253	249 + 253 + 255 = 757	757/3 = 252.33
1955	255	253 + 255 + 251 = 759	759/3 = 253.00
1956	251	255 + 251 + 257 = 763	763/3 = 254.33
1957	257	251 + 257 + 260 = 768	768/3 = 256.00
1958	260	257 + 260 + 265 = 782	782/3 = 260.67
1959	265	260 + 265 + 262 = 787	787/3 = 262.33
1960	262	–	–

6. Obtain the straight line trend equation for the following data by the method of the least square. Tabulate the trend values.

Year :	2010	2011	2012	2013	2014	2015	2016
Sale (in '000 units)	140	144	160	152	168	176	180

Sol : **(Jan.-21)**

$$Y_c = a + bx$$

$$a = \frac{\Sigma y}{n} \quad b = \frac{\Sigma xy}{\Sigma x^2}$$

Fitting of straight line trend

Year	Sales	x	xy	x^2	$y_c = a + bx$
2010	140	– 3	– 420	9	139.42
2011	144	– 2	– 288	4	146.28
2012	160	– 1	– 160	1	153.14
2013	152	0	0	0	160
2014	168	1	168	1	166.86
2015	176	2	352	4	173.72
2016	180	3	540	9	180.58
	1120		192	28	

$$a = \frac{1120}{7} = 160$$

$$b = \frac{192}{28} = 6.86$$

$y_c = 160 + 6.86(x)$

For 2010 = 160 + 6.86(– 3)
160 – 20.58
= 139.42

For 2011 = 160 + 6.86(– 2)
160 – 13.72
= 146.28

For 2012 = 160 + 6.86(– 1)
160 – 6.86
= 153.14

For 2013 = 160 + 6.86(0)
160 – 0
= 160

For 2014 = 160 + 6.86(1)
160 + 6.86
= 166.86

For 2015 = 160 + 6.86(2)
160 + 13.72
= 173.72

For 2016 = 160 + 6.86(3)
= 160 + 20.58
= 180.58

7. Fit a straight line by the Least Square Method and tabulate the trend values for the above data.

Year	2011	2012	2013	2014	2015	2016	2017
Production (in tons)	77	88	94	85	91	98	90

Sol : **(June-19, Imp.)**

Fitting of Straight line trend

Years	Y	Years – Middle Year = X Years – 2014 = X	X^2	XY	Y_c
2011	77	– 3	9	– 231	83
2012	88	– 2	4	– 176	85
2013	94	– 1	1	– 94	87
2014	85	0	0	0	89
2015	91	+ 1	1	91	91
2016	98	+ 2	4	196	93
2017	90	+ 3	9	270	95
	623		28	56	623

$$a = \frac{\Sigma y}{N} = \frac{623}{7} = 89$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{56}{28} = 2$$

Calculation of Trend Values (Y_c)

2011 $Y_c = a + bx = 89 + [2]\,[-3] = 89 - 6 = 83$

2012 $Y_c = a + bx = 89 + [2]\,[-2] = 89 - 4 = 85$

2013 $Y_c = a + bx = 89 + [2]\,[-1] = 89 - 2 = 87$

2014 $Y_c = a + bx = 89 + [2]\,[0] = 89 + 0 = 89$

2015 $Y_c = a + bx = 89 + [2]\,[1] = 89 + 2 = 91$

2016 $Y_c = a + bx = 89 + [2]\,[2] = 89 + 4 = 93$

2017 $Y_c = a + bx = 89 + [2]\,[3] = 89 + 6 = 95$

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 : **(Nov.-21)**

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 $= \frac{18}{82.6} \times 141.5 = 30.83$

9. 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 : **(Aug.-21)**

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

5.4 Index Numbers

5.4.1 Introduction

Q13. Define Index Numbers.

(OR)

Give a brief introduction of index numbers.

(OR)

What is an index numbers?

Ans : **(May-22, Aug.-21)**

Meaning

An Index, simply stated, is an indicator. It indicates the broad change in a given phenomenon. It is relative and indicates changes in the level of the given phenomenon in course of time, across geographical locations or in respect of any other characteristic. For example, the BSE Sensex (stands for Bombay Stock Exchange Sensitive Index) is an index of movement in stock prices. It tells us, very broadly, whether the prices of stocks of various companies have gone up or moved down on the Bombay Stock Exchange. Similarly, the Wholesale Price Index (WPI) gives us an indication of whether the general price level in the economy is going up or falling down. Similarly, Index numbers (or indices) can be constructed to measure any phenomenon, such as Industrial production, Number of road accidents, cost of Living or any other activity. Thus, Index Numbers are barometers measuring change in the level of a phenomenon.

Definitions

(i) According to Croton and Cowden Different Experts have defined Index Numbers in different words. Some of the definitions are stated as below.

(ii) According to Speiges "Index numbers are devices for measuring differences in the magnitude of a group of related variables."

(iii) According to Weldon "An Index Number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographical location or any the characteristic."

(iv) According to Edgeworth "An Index Number is a statistical device for indicating the relative movements of data where measurement of actual movement is difficult or incapable of being made."

(v) According to Bowley "Index Number shows by its variations the changes in magnitude which is not susceptible either of accurate measurement in itself or of direct valuation in practice."

"Index Numbers are used to measure the change in some quantity which cannot be observed directly, which we know to have a definite influence on many other quantities which we can so observe, tending to increase all or diminish all, while this influence is concealed by the action of many causes affecting the separate quantities in various ways."

5.4.2 Characteristics and Uses of Index Numbers

Q14. Explain the Characteristics of Index Numbers.

Ans :

On the basis of the above definitions, the following points can be made:

1. Index Numbers are a measurement device
2. They do not measure or state the actual level attained by the phenomenon being studied. They measure the change in the phenomenon being studied.
3. The situations for which Index Numbers are used for comparison are not restricted in any manner. It can be a comparison of two time periods, two geographical locations, two groups of people or any other phenomenon.
4. Index Numbers are the result of a numerical calculation. They do not have any units such as kgs or rupees.
5. Index Numbers are relative terms and hence, they are normally expressed in percentage terms.

Q15. What are the uses of index numbers ?

(OR)

Explain the importance of index numbers.

Ans : **(May-22)**

1. Economic Barometers

Index Numbers can be constructed for any phenomenon for which quantitative information is available. They capture the various changes taking place in the general economy and business activities. They provide a fair view of the general trade, the economic development and business activity of the country. Thus, they are aptly termed as 'Economic barometers'.

2. Study of Trend

Index Numbers study the relative changes in the level of a phenomenon over different periods of time. They are especially useful for study of general trend of the general trend for a group phenomenon in a time series data.

3. Policy Formulation

Index Numbers are indispensable for any organization in efficient planning and formulation of executive decisions. For example, the Dearness Allowance payable to employees is determined on the basis of Cost of living index numbers. Similarly, Psychiatrists use Intelligence Quotients to assess a child's intelligence in relation to his/her age, which further can be used in framing the education policy.

4. Deflation

Index Numbers can be deflated to find out the real picture pertaining to the phenomenon. For example, we can know if the real income of employees has been increasing by deflating the nominal wages with the help of index numbers.

5. Forecasting

Index Numbers provide valuable information that aids forecasting. For example an Index of sales, along with related indices such as cost of living index, helps in forecasting future demand and future sales of business.

6. Measurement of Purchasing Power

The cost of Living Index helps us in finding out the intrinsic worth of money. It is one of the key indicators touching the life of the common man.

7. Simplicity

Index Numbers eliminate the clutter of large numbers and complicated calculations, providing the underlying information in a manner that is simple and easy to underlying information in a manner that is simple and easy to understand. For example, many people may not understand the dynamics of stock markets, but they can follow the movements of the BSE Sensex with ease.z

---

Q16. Describe briefly applications of index numbers in business and industry.

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

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.

5.5 Types of Index Numbers

5.5.1 Unweighted Price Indexes, Weighted Price Indexes

Q17. What are the various methods of Constructing Index Numbers?

Ans : **(Imp.)**

The various methods of constructing index numbers are shown in the following figure :

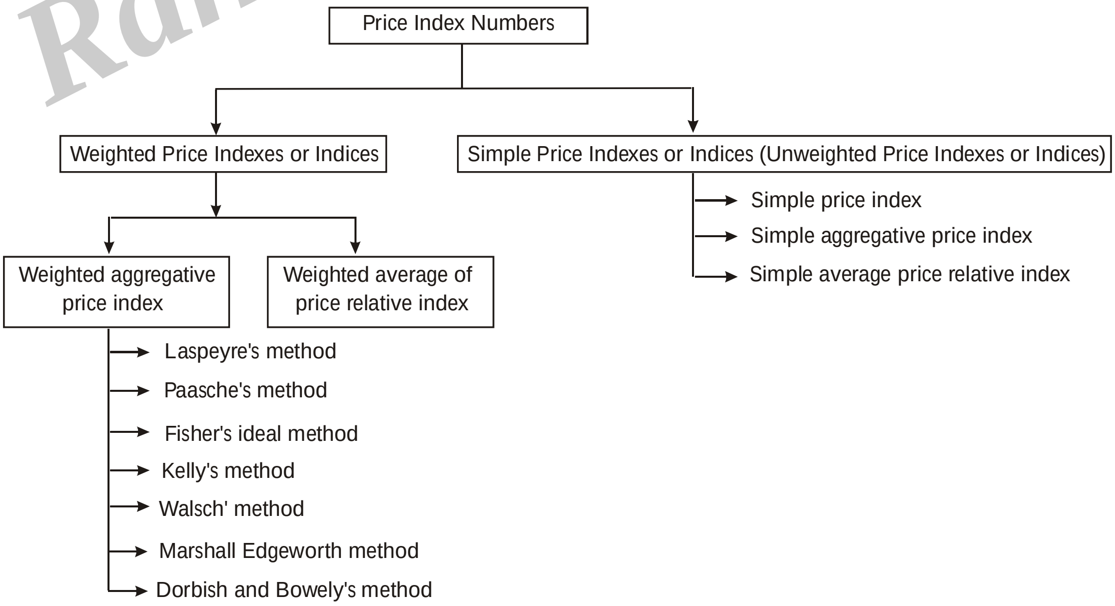


Fig : Methods of Price Index Numbers

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) Paasche'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 Relatives Method

The basis methodology of calculating the price relatives is same as in case of simple average of relatives method. However, instead of calculating simple average weights are assigned to price relatives and a weighted average is calculated. Such weighted average can be arithmetic mean or geometric mean of the weighted price relatives. The following steps have to be followed.

Case (i) If weighted Arithmetic Mean is used.

Step 1:

Calculate Price Relatives P for each item, $p = (P, /P_0) \times 100$

Step 2:

Calculate weights with which price relatives are to be multiplied.Normally, the value of the item in the base year (i.e. P_0q_0) is taken as weight. How ever, weights can be P_0q, p_1q_0 or p_1q_1. Weights are denoted by V.

Step 3 :

Calculate weighted price relatives by multiplying the price relatives with their corresponding weights. In other words, Calculate PV.

Step 4 :

Add weighted price relatives obtained in step 3. Denote as $\sum PV$

Step 5 :

Calculate the sum of weights. This is denoted by $\sum V$

Step 6 :

Weighted Average of Price Relatives $= p_{0.1} = \frac{\Sigma PV}{\Sigma V}$

Case (ii) If Geometric mean is to be used.

Step 1 :

Calculate Price relatives P for each item. $P = (P_1/p_0) \times 100$

Step 2 :

Calculate the logarithm value of P. This is denoted by log P

Step 3 :

Calculate weights for each them. This is denoted by V.

Step 4 :

Calculate V log P for each item by multiplying weight obtained in step 3 with logarithm value of price relatives obtained in step 2.

Step 5 :

Add the weighted logarithm values of Price relatives. Denote it as $\sum(V.\log P)$.

Step 6 :

Weighted Average to price relatives

$$= P_{0.1} = \text{Antilog}\left[\frac{\sum(V.\log P)}{\sum V}\right]$$

2. Unweighted Price Index

(a) Simple Aggregative Method

This method involves aggregation of prices in the current period and expressing the total as a percentage of aggregate of prices in the base period. The following steps are followed.

Step 1:

Calculate $\sum P_1, \sum P_1$ is the sum total of prices of all items in the current year.

Step 2:

Calculate $\sum P_0, \sum P_0$ is the sum total of prices of all items in the base year.

Step 3:

Calculate $P_{0.1}$, $P_{0.1}$, is the price index number of the current year with respect to the base year. It is expressed in percentage terms and calculated as under:

$$P_{0.1} = \left(\frac{\sum P_1}{\sum P_0}\right) \times 100$$

(b) Simple Average of Price Relative Method

Under this method, the price of each item in the current year is expressed as a percentage of its price in the base year. The figure so obtained is called Price Relative. The price relatives are then averaged to calculate the index number for that year. Either Arithmetic mean or Geometric mean can be used for calculation of average of price relatives. The following steps are following.

(a) If Arithmetic mean is used for the purpose of averaging then

Step 1:

Calculate Price relative for each item. Price Relative of an item. Is obtained by the formula $\left(\frac{p_1}{p_0}\right) \times$ 100

Where P1 is price of the item in the current year and

P0 is the price of the item in the base year.

Step 2 :

Calculate average of price relatives to obtain the Index Number P0.1

$$P_0.1 = \sum \frac{\left(\frac{p_1}{p_0} \times 100\right)}{N}$$

Where N = Number of Items

(b) If Geometric Mean is used for averaging the price relatives, then

Step 1 :

Calculate Price Relative (p) of each item (p = P1/ P_0) × 100

Step 2 :

Calculate Logarithm Value of each Price relative (log p)

Step 3 :

Calculate simple average of logarithm values obtained in step $2\left(\frac{\text{Logp}}{N}\right)$

Step 4 :

Calculate Antilog of value obtained in step 3 Thus

$$\text{Index Number } P_{0.1}, = \text{Antilog}\left[\frac{\sum \log p}{N}\right]$$

$$\text{Where } p = \left(\frac{p_1}{p_0} \times 100\right)$$

Q18. What is Laspeyre's index method ?

Ans :

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 1:

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 $\sum p_1q_0$

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 $\sum p_1q_0$

Step 5:

The index number as per Laspeyre's method

$$= P_{0.1} = \frac{\sum P_1q_0}{\sum P_0q_0} \times 100$$

Q19. What is Paasche's index method ?

Ans :

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 $\sum p_1q_1$

Step 3 :

Multiply base year's Prices (p0) with current year's quantities (q_1)

Step 4 :

Add the numbers obtained in step (2). The resultant sum is $\sum p_0q_1$

Step 5 :

Index Number as per Paasche's method $= P_{0.1} = \dfrac{\sum p_1q_1}{\sum p_0q_1} \times 100$

Q20. Compare and Constrat Laspeyre's and Paasche's method ?

Ans :

S.No.	Laspeyre's Index Number	S.No.	Paasche's Index Number
1.	Here, quantity of the base year is assumed to be the quantity of the current year.	1.	Here, quantity of the current year is assumed to be the quantity of the base year.
2.	It has an upward bias i.e. the numerator of the index number is increased due to the assignment of higher weights fixed on the basis of the base year's quantities even though there might have been a fall in the quantity consumed during the current year due to rise, or fall in price and change in in tastes, habits and customs etc. in the current year.	2.	It has a downward bias i.e. the numerator of the index number is decreased due to the assignment of lower weights fixed on the basis of the current year's quantities even though the quantities in the current year might have fallen due to rise or fall in price, or change in habits of consumption.
3.	As the quantity of the base year are used as weights, the influence of price changes on quantities demanded do not get reflected in the index number.	3.	As the quantities of the current year are used as weights, the influence of price changes on quantities demanded get reflected in the index number.
4.	It measures changes in a fixed marked basketof goods and services as the same quantities are used in each period.	4.	It continually updates the quantities to the level of current consumption.
5.	Here, weights remain constant.	5.	Here, weights are determined every time an index number is constructed.

Q21. What is Marshall Edgeworth?

Ans :

In this method also both the current year as well as base year prices and quantities are considered. The formula for constructing the Index is :

$$P_{01} = \frac{\Sigma(q_0 + q_1)p_1}{\Sigma(q_0 + q_1)p_0} \times 100$$

on opening the brackets

$$P_{01} = \frac{\Sigma\, p_1q_0 + \Sigma\, p_1q_1}{\Sigma\, p_0q_0 + \Sigma\, p_0q_1} \times 100$$

It is a simple, readily constructed measure, giving a very close approximation to the results obtained by the ideal formula.

Q22. What is 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.

$$P_{0.1} = \left[\sqrt{\frac{\sum p_1q_0 \times \sum p_1q_1}{\sum p_0q_0 \times \sum p_0q_1}}\right] \times 100$$

Reasons for Fisher's Index being called an Ideal Index

Reasons

(i) It gives weightage to both current consumption and base year consumption.

(ii) It is free from upward or downward bias.

(iii) It satisfies both time reversal and factor reversal tests (to be discussed later) .

(iv) It is a Geometric mean of Laspeyre's index and Paasche's index

Q23. What are the problems involved in construction of index numbers ? Explain.

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

The following problems are mainly faced in the construction of index numbers :

i) Definition of the Purpose,

ii) Selection of the Base Period,

iii) Selection of Items,

iv) Selection of Sources of Data and Collection of Data,

v) Selection of Average,

vi) System of Weighting.

(i) Definition of the Purpose

There are no all purpose index numbers. Therefore, before constructing an index number the specific purpose, i.e., objective for which it is designed must be clearly and rigorously defined. Haberler has rightly said "Different index numbers are constructed to fulfil different objectives and before setting to construct a particular number one must clearly define one's objective of study because it is on the objective of the study that the nature and format of the index number depends".

(ii) Selection of the Base Period

Selection of the proper Base Period is an important factor in the construction of index numbers. Base Period is a reference point with which changes in other periods are measured. About the selection of Base Period, following observations may be noted.

Morris Hamburg observes, "It is desirable that the Base Period be not too far away in times from the present. The further away we move from the Base Period, the dimmer are our recollections of Economic conditions prevailing at that time. Consequently comparisons with these remote periods tend to lose significance and to become rather tenuous in meaning."

According to George Simpson and Fritz Katka, "Since practical decisions are made in terms of index numbers, and economic practices so often are a matter of the short run, we wish to make comparisons between a base which lies in the same general economic framework as the years of immediate interest. Therefore, we choose a base relatively close to the years being studied." In the fast changing world of today the base year should not be more than a decade old. There is a psychological reason also for taking a recent period as base.

(iii) Selection of Items

Selection of items or 'Regimen' or 'Basket'. In any index number neither it is possible nor necessary to include all the items or commodities. Each index number tries to measure changes pertaining to a particular group.

Selection of items also depends on the purpose of index number. Moreover, items selected should be such as are widely consumed. The items selected for an index number should be relevant, representative, reliable and comparable. In general the larger is the number of items, the lesser will be the chances of error in the average. Since we cannot include a very large number of items. A compromise is always required between the number of items and the reasonable standard of accuracy. We must, however, have manageable number of items and should also aim at reasonable standard of accuracy.

(iv) Selection of Sources of Data and Collection of Data

For sources of data and collection of data we are mainly concerned with she prices. Use of wholesale prices or retail prices depends on the objective of study. Price quotations should be obtained from important markets. In order to ensure better results, it is advisable to take a standard price which implies representative price of a commodity for whole interval under consideration.

(v) Selection of Average

Since index numbers measure the relative changes, Geometric mean should be the best average but due to certain difficulties in calculations with G.M., for all practical purposes Arithmetic mean is used.

(vi) System of Weighting

According to John Giffin, "In simple terms, weighting is designed to give component series an importance in proper relation to their real significance." In order to allow each commodity to have a reasonable influence on the index it is advisable to use a suitable weighting system.

In case of an unweighted index number of prices, all commodities are given equal importance. But in actual practice different commodities need a different degree of importance.

The weights may be according to :

- The value or quantity produced.
- The value or quantity consumed.
- The value or quality sold.

PROBLEMS

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

Month	Price Index	New price Index
June - 2018	142	$\frac{142}{144}\times 100$ = 98.61
July - 2018	137	$\frac{137}{144}\times 100$ = 95.13
August - 2018	143	$\frac{143}{144}\times 100$ = 98.30
September - 2018	142	$\frac{142}{144}\times 100$ = 98.61
October - 2018	145	$\frac{145}{144}\times 100$ = 100.69
November - 2018	151	$\frac{151}{144}\times 100$ = 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

11. From the following data given below compute Fisher's ideal index for the year 2017

Items	2013		2017	
	Price	Quantity	Price	Quantity
P	4	74	6	82
Q	10	125	8	140
R	14	40	12	33

Sol :

Computation of the Laspeyre's Index number

for 2017 with 2014 as the base year

Items	2013		2017					
	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
P	4	74	6	82	444	296	492	328
Q	10	125	8	140	1000	1250	1120	1400
R	14	40	12	33	480	560	396	462
					1924	2106	2008	2190

Marshall and Endgeworth's index number is given by

$$P_{01}(ME) = \frac{\sum p_1q_0 + \sum p_1q_1}{\sum p_0q_0 + \sum p_0q_1} \times 100 = \frac{1924 + 2008}{2106 + 2190} \times 100 = \frac{3932}{4296} \times 100$$

$$= 91.53 \text{ approx.}$$

Fisher's idel Index is given by

$$P_{01}(F) = \sqrt{\frac{\sum p_1p_0}{\sum p_0p_1} \times \frac{\sum p_1q_1}{\sum p_0q_1}} \times 100 = \sqrt{\frac{1924}{2106} \times \frac{2008}{2190}} \times 100$$

$$= \sqrt{0.9135 \times 0.9168} \times 100 = \sqrt{0.83749} \times 100 = 0.9151 \times 91.51$$

From the above results of the two indices, it is clear that the index number of Marshall and Edgeworth is a close approximation of the ideal index of Fisher.

12. Construct index numbers of price from the following data by applying :

(i) Laspeyres method

(ii) Paasche method

(iii) Fisher's ideal method, and

(iv) Marshall-Edgeworth method

Commodity	2010		2011	
	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	2010		2011		$p_1 q_0$	$p_0 q_0$	$p_1 q_1$	$p_0 q_1$
	Price (P_0)	Qty (Q)	Price (P_1)	Qty (q_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	58	50	40
D	2	19	2	13	38	38	26	26
					200	160	130	103

(i) Laspeyres Method : $p_{01} = \frac{\Sigma\, p_1q_0}{\Sigma\, p_0q_0} \times 100$; where $\Sigma p_1q_0 = 200.\ \Sigma p_0q_0 = 160$

$$p_{01} = \frac{200}{160} \times 100 = 125$$

(ii) Paasche's Method : $p_{01} = \frac{\Sigma\, p_1q_1}{\Sigma\, p_0q_1} \times 100$; where $\Sigma p_1q_1 = 130.\ \Sigma p_0q_1 = 103$

$$p_{01} = \frac{130}{103} \times 100 = 126.21$$

(iii) Fisher's Ideal Method : $p_{01} = \sqrt{\frac{\Sigma\, p_1q_0}{\Sigma\, p_0q_0} + \frac{\Sigma\, p_1q_1}{\Sigma\, p_0q_1}} \times 100 = \sqrt{\frac{200}{160} + \frac{130}{103}} \times 100$

$$= \sqrt{1.578} \times 100 = 1.256 \times 100 = 125.6$$

(iv) Marshall-Edgeworth Method : $p_{01} = \frac{\Sigma(q_0 + q_1)p_1}{\Sigma(q_0 + q_1)p_0} \times 100 = \frac{\Sigma p_1q_0 + \Sigma p_1q_1}{\Sigma p_0q_0 + \Sigma p_0q_1}$

$$= \frac{200 + 130}{160 + 103} \times 100 = \frac{330}{363} \times 100 = 125.47$$

13. From the following data calculate price index according to

(i) Laspeyre,

(ii) Paasche and

(iii) Marshall-Edgeworth methods.

Item	Base year		Current year	
	Price (`)	Expenditure (`)	Price (`)	Expenditure (`)
A	5	50	8	40
B	7	25	12	30
C	9	10	15	25
D	12	5	20	18

Sol : **(Jan.-21)**

$$\text{Qty} = \frac{\text{Expenditure}}{\text{Price}}$$

Base year

A $\frac{50}{5} = 10$

B $\frac{25}{7} = 3.6$

C $\frac{10}{9} = 1.1$

D $\frac{5}{12} = 0.42$

Base year

A $\frac{40}{8} = 5$

B $\frac{30}{12} = 2.5$

C $\frac{25}{15} = 1.7$

D $\frac{18}{20} = 0.9$

Item	p_0	q_0	p_1	q_1	p_1q_1	p_1q_0	p_0q_1	p_0q_0
A	5	10	8	5	40	80	25	50
B	7	3.6	12	2.5	30	43.2	17.5	25.2
C	9	1.1	15	1.7	25.5	16.5	15.3	9.9
D	12	0.42	20	0.9	18	8.4	10.8	5.04
					113.5	148.1	68.6	90.14

(i) Laspeyre method $= p_{01} = \dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$

$$= \frac{14.81}{90.14} \times 100 = 164.29$$

(ii) Passche method $= p_{01} = \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

$$= \frac{113.5}{68.6} \times 100 = 165.45$$

(iii) Marshall Edgeworth

$$p_{01} = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times 100$$

$$= \frac{148.1 + 113.5}{90.14 + 68.6} \times 100$$

$$= \frac{261.6}{158.74} \times 100 = 164.79$$

14. Give a brief introduction of Index Numbers What are their uses?

Construct Laspayre and Paasche Index Numbers for the following data:

Commodity	2020 Price	2020 Quantity	2021 Price	2021 Quantity
P	3	10	5	7
Q	6	12	7	6
R	5	16	6	10
S	3	20	3	12

Sol : **(May-22)**

Commodity	P_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
P	3	10	5	7	50	30	35	21
Q	6	12	7	6	84	72	42	36
R	5	16	6	10	96	80	60	50
S	3	20	3	12	60	60	36	36
					290	242	173	143

(i) Laspeyres (P_0 1) $= \dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$

$$= \frac{290}{242} \times 100 = 119.83$$

(ii) Paaschs (P_0, 1) $= \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

$$= \frac{173}{143} \times 100 = 120.97$$

5.5.2 Tests of Adequacy

Q24. 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

15. 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 :

Commodity	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 :

$$P_{0,1} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}} \times 100$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92}} \times 100$$

$$= \sqrt{\frac{40{,}198}{8372}} \times 100 = \sqrt{4.8015} \times 100$$

$$= 2.191 \times 100 = 219.1$$

Time Reversal Test :

$$P_{0,1} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}}$$

$$P_{1,0} = \sqrt{\frac{\sum p_0q_1}{\sum p_1q_1} \times \frac{\sum p_0q_0}{\sum p_1q_0}}$$

$$P_{0,1} \times P_{1,0} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1} \times \frac{\sum p_0q_1}{\sum p_1q_1} \times \frac{\sum p_0q_0}{\sum p_1q_0}}$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92} \times \frac{92}{199} \times \frac{91}{202}} = \sqrt{1} = 1$$

$$\therefore P_{0,1} \times P_{1,0} = 1$$

Factor Reversal Test

$$P_{0,1} \times q_{0,1} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1} \times \frac{\sum p_0q_1}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_1q_0}}$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92} \times \frac{92}{91} \times \frac{199}{202}} = \sqrt{\frac{199}{91} \times \frac{199}{91}} = \frac{199}{91}$$

$$V_{0,1} = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} = \frac{199}{91}$$

$$\therefore P_{0,1} \times Q_{0,1} = V_{0,1}$$

16. 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 $= P_{0,1} = \sqrt{\dfrac{\Sigma p_1q_o}{\Sigma p_oq_o} \times \dfrac{\Sigma p_1q_1}{\Sigma p_oq_1}} \times 100$

$$= \sqrt{\frac{3,365}{3,530} \times \frac{3,400}{3,600}} \times 100$$

$$= \sqrt{0.95 \times 0.094} \times 100$$

$$= \sqrt{0.0893} \times 100 = 0.298 \times 100$$

$$= 29.8$$

Time Reversal Test $= P_{0,1} \times P_{1,0} = 1$

$$= \sqrt{\frac{\Sigma p_1q_o}{\Sigma p_oq_o} \times \frac{\Sigma p_1q_1}{\Sigma p_oq_1}} \times \sqrt{\frac{\Sigma p_oq_1}{\Sigma p_1q_1} \times \frac{\Sigma p_oq_o}{\Sigma p_1q_o}}$$

$$= \sqrt{\frac{3,3\cancel{6}5}{3,5\cancel{3}0} \times \frac{\cancel{3},400}{3,\cancel{6}00}} \times \sqrt{\frac{3,6\cancel{0}0}{3,\cancel{4}00} \times \frac{3,5\cancel{3}0}{3,3\cancel{6}5}}$$

$$= \sqrt{1} = 1$$

Hence Time Reversal Test is Satisfied.

17. 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 $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$. Hence, the above data also satisfies the Factor Reversal Test.

18. 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 : $P_{0,1} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$

$$= \sqrt{\frac{2015}{1774} \times \frac{2172}{1914}} \times 100 = 1.135 \times 100 = 113.5$$

Time Reversal Test : Time reversal test is satisfied when :

$$P_{0,1} \times P_{1,0} = 1$$

$$P_{1,0} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} = \sqrt{\frac{1914}{2172} \times \frac{1774}{2015}}$$

$$P_{0,1} \times P_{1,0} = \sqrt{\frac{2015}{1774} \times \frac{2172}{1914} \times \frac{1914}{2172} \times \frac{1774}{2015}}$$

Hence, time reversal test is satisfied by the given data.

Factor Reversal Test : Factor reversal test is satisfied when :

$$P_{0,1} \times q_{0,1} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

$$q_{0,1} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} = \sqrt{\frac{1914}{1774} \times \frac{2172}{2015}}$$

$$P_{0,1} \times q_{0,1} = \sqrt{\frac{2015}{1774} \times \frac{2172}{1914} \times \frac{1914}{1774} \times \frac{2172}{2015}} = \frac{2172}{1774}$$

$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$ is also equal to $\frac{2172}{1774}$. Hence factor reversal test is satisfied by the given data.

5.6 Consumer Price Indexes

Q25. Discuss in detail consumer price index numbers/cost of living index numbers.

Ans :

Meaning

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

1. Consumer Price Index Numbers (CPIN) are used in the preparation of wage contracts and wage negotiations.
2. CPIN assists the government and business organization 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.

Q26. 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 :

$$\text{Consumer Price Index} = \text{Cost of Living Index} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_1} \times 100$$

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. $\sum p_0 q_0$ is total expenditure incurred in the base period. $\sum p_1 q_0$ 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

19. 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 :

$$CP_1 = \frac{\sum PV}{\sum V} = \frac{370250}{2605} = 142.13$$

20. 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 $= \text{Antilog}\left[\frac{\sum W \log I}{\sum W}\right] = \text{anti}\log\left[\frac{24.4016}{10}\right]$

$= \text{Antilog} \ (2.44016) = 275.55$

5.7 Report Writing

5.7.1 Importance of Report Writing

Q27. Define Report, Report Writing and Research Report. State the importance of Report Writing.

Ans : **(May-22, 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

Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented and/or written. As a matter of fact even the most brilliant hypothesis, highly well designed and conducted research study, and the most striking generalizations and findings are of little value unless they are effectively communicated to others.

The purpose of research is not well served unless the findings are made known to others. Research results must invariably enter the general store of knowledge. All this explains the significance of writing research report. There are people who do not consider writing of report as an integral part of the research process.

But the general opinion is in favour of treating the presentation of research results or the wiiting of report as part and parcel of the research project. Writing of report is the last step in a research study and requires a set of skills somewhat different from those called for in respect of the earlier stages of research. This task should be accomplished by the researcher with utmost care; he may seek the assistance and guidance of experts for the purpose.

(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

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.

5.7.2 Types of Research Reports

Q28. 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 organizations 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.

5.7.3 Report Preparation and Presentation

Q29. Explain the contents of report.

(OR)

Explain the process of Report Preparation and Presentation.

Ans : **(Imp.)**

The researcher must keep in mind that his research report must contain following aspects:

1. Purpose of study
2. Significance of his study or statement of the problem
3. Review of lit era true
4. Methodology
5. Interpretation of data
6. Conclusions and suggestions
7. Bibliography
8. Appendices

These can be discussed in detail as under:

1. Purpose of Study

Research is one direction-oriented study. He should discuss the problem of his study. He must give background of the problem He must lay down his hypothesis of the study. Hypothesis is the statement indicating the nature of the problem He should be able to collect data, analyze it and prove the hypothesis. The importance of the problem for the advancement of knowledge or removal of some evil may also be explained. He must use review of literature or the data from secondary source for explaining the statement of the problems.

2. Significance of Study

Research is re-search and hence the researcher may highlight the earlier research in new manner or establish new theory. He must refer earlier research work and distinguish his own research from earlier work. He must explain how his research is different and how his research topic is different and how his research topic is important. In a statement of his problem, he must be able to explain in brief the historical account of the topic and way in which he can make and attempt. In lus study to conduct the research on his topic.

Purpose of Study
↓
Significance of Study
↓
Review of Literature
↓
Methodology
↓
Interpretation of Data
↓
Conclusions/Suggestions
↓
Bibliography
↓
Appendices

Fig.: Process of Report Preparation

3. Review of Literature

Research is a continuous process. He cannot avoid earlier research work. He must start with earlier work. He should note down all such research work, published] in books, journals or unpublished thesis. He will get guidelines for his research from taking a review of literature. He should collect information in respect of earlier research work. He should enlist them in the manner given below:

(i) Author Researcher

(ii) Title of research / Name of book

(iii) Publisher

(iv) Year of publication

(v) Objective of his study

(vi) Conclusion Suggesuons

Then he can compare this information with his study to show separate identity' of his study. He must be honest to point out similarities and differences of his study from earlier research weak.

4. Methodology

It is related to colleen on of data. There are two sources for collecting data primary and secondary. Primary data is original and collected m field weak, either through questionnaire' interviews. The secondary' data relied on library work. Such primary data are collected by sampling method The procedure for selecting the sample must be mentioned The methodology-must give various aspects of the problem that are studied for valid generalization about the phenomena. The scales of measurement must be explained along with different concepts used in the study.

While conducting a research based on field weak, the procedural things like definition of universe, preparation of source list must be given We use case study method historical research, etc. He must make it clear as to which method is used in his research work. When questionnaire is prepared, a copy of it must be given in appendix.

5. Interpretation of Data

Mainly the data collected from primary source need to be interpreted m a systematic manner. The tabulation must be completed to draw conclusions. All the questions are not useful for report writing.

One has to select them or dub them according to hypothesis or objectives of study.

6. Conclusions/Suggestions

Data analysis forms the crux of the problem. The information collected in field work is useful to draw conclusions of study. In relation with the objectives of study the analysis of data may lead the researcher to pin point his suggestions. This is the most important part of study. The conclusions must be based on logical and statistical reasoning. The report should contain not only the generalization of inference but also the basis on which the inferences are drawn. All sots of proofs, numerical and logical must be given m support of any theory that has been advanced He should point out the limitations of his study.

7. Bibliography

The list of references must be arranged in alphabetical order and be presented in appendix. The books should be given in first section and articles are m second section and research projects in the thud. The pattern of bibliography is considered convenient and satisfactory from the point of view of reader.

8. Appendices

The general information in tabular form which is not directly used in the analysis of data but which is useful to understand the background of study can be given in appendix.

Q30. 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. Organization of Presentation

Clarity in presentation is essential that comes with proper organization 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 organization of presentation enhances the effectiveness of the presentation. On the other hand, improper organization 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.

5.7.4 Report Structure

Q31. Describe the structure of report.

Ans :

There is scientific method for the layout of the research report. The layout of the report means as to what the research report should contain. The contents of the research report are noted below:

1. Preliminary Page
2. Main Text
3. End Matter

1. Preliminary Pages

These must be title of the research topic and data. There must be preface or foreword to die research work. It should be followed by table of contents. The list of tables, maps should be given

2. Main Text

It provides the complete outline of research report along with all details. The title page is reported in the main text. Details of text are given continuously as divided in different chapters.

(a) Introduction

(b) Statement of the problem

(c) The analysis of data

(d) The implications drawn from the results

(e) The summary

(a) Introduction: Its purpose is to introduce the research topic to readers. It must cover statement of the problem, hypotheses, objectives of study, review of literature, and the methodology to cover primary and secondary data, limitations of study and chapter scheme. Some may give in brief m the first chapter the introduction of the research project highlighting the importance of study. This is followed by research methodology in separate chapter.

The methodology should point out the method of study, the research design and method of data collection.

(b) **Statement of the Problem:** This is crux of his research. It highlights main theme of his study. It must be in non technical language. It should be in simple manner so ordinary reader may follow it. The social research must be made available to common man The research in agricultural problems must be easy for farmers to read it.

(c) **Analysis of Data:** Data so collected should be presented in systematic manner and with its help, conclusions can be drawn. This helps to test the hypothesis. Data analysis must be made to confirm the objectives of the study.

(d) **Implications of Data:** The results based on the analysis of data must be valid. This is the main body of research. It contains statistical summaries and analysis of data. There should be logical sequence in the analysis of data. The primary data may lead to establish the results. He must have separate chapter on conclusions and recommendations. The conclusions must be based on data analysis. The conclusions must be such which may lead to generalization and its applicability in similar circumstances. The conditions of research work limiting its scope for generalization must be made clear by the researcher.

(e) **Summary:** This is conclusive part of study. It makes the reader to understand by reading summary the knowledge of the research work. This is also a synopsis of study.

3. End Matter

It covers relevant appendices covering general information, the concepts and bibliography. The index may also be added to the report.

5.7.5 Report Formulation

Q32. 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:

(i) Bio-data of promoter

Name and address, qualifications, experience and other capabilities of the entrepreneur. Similar information of each partner if any.

(ii) Industry profile

A reference analysis of industry to which the project belongs, e.g., past performance; present status, its organization, its problems etc.

(iii) 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.

(iv) 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.

(i) 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.

(ii) 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.

Q33. What are the purpose of footnotes? What is the kind of footnotes one finds in research?

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

Meaning

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.

5.7.6 Guides for Effective Documentation

Q34. What is Documentation. Explain the Guides for Effective Documentation.

Ans : **(Imp.)**

Meaning

Documentation is the process of collecting and extracting the documents which are relevant to research.

Documents may be classified into:

i) Personal documents

ii) Company documents

iii) Consultants report and published materials: and

iv) Public documents

(i) Personal documents

Personal documents are those that are written by or on behalf of individuals. They may include autobiographies, biographies, diaries, memoirs, letters, observations and inscriptions, which are primarily written for the use and satisfaction of individuals and which can be utilized for research purposes. Personal documents play a very vital role in research, when certain information is not available from any other sources: however, such documents are subject to the difficulties of availability, reliability and validity of inferences.

(ii) Company documents

Company documents are the most essential types of documents is management research, annual reports, statements of income and expenditure and balance sheets, files and records, policy statements, resolutions, minutes of board of directors, general bodies and executive conferences, performance records and evaluation files, specific forecasting and evaluation reports, directors' reports, etc. Many of these documents are reliable ones, though they are subject to the problems of availability and adequacy. Such documents can be published or unpublished.

(iii) Consultants report and published materials

Consultant's published materials consist of report of professional consultants, records of commodity boards, chambers of commerce. FICCL manufacturers' associations and industry associations.

(iv) Public documents

Public documents are documents, both published and unpublished, of government organizations and documents of pubic interest. They include government records and files, draft outlines of five-year plans, consultative committee reports, finance commission reports, special enquiry' commission reports. Company Law Board reports. MRTP Commission reports, reports and files of the Registrar of Companies, the Registrar of Firms, the Ministry of Commerce and Industry, etc., report of population census. National Sample Survey and such other government research institutions. Such documents are valuable if they are reliable and suitable for a particular study if they can be obtained. Documentation is one of the most important needs of any management researcher at the primary sate of his research.

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.

Q35. What care must be taken in designing capitalization in report preparation?

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

- 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.

5.7.7 Research Briefings (Oral Presentation)

Q36. 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 organizations are usually presented in the form of technical reports.

5.7.8 Referencing Styles and Citation in Business Management Research

Q37. 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.

Short Question and Answers

1. Define time series.

Ans :

Meaning

A time series is a statistical data that are collected, observed or recorded at regular intervals of time. The term time series applies, for example, to the data recorded periodically showing the total annual sales of retail stores, the total quarterly value of construction contracts awarded, the total amount of unfilled orders in durable goods industries at the end of each month, weekly earnings of workers in an industrial town, hourly temperature in a particular city.

Definitions

Some of the important definitions of time series, given by different experts are as under:

(i) According to Morris Hamburg, "A time series is a set of statistical observations arranged in chronological order."

(ii) According to Patterson, "A time series consists of statistical data which are collected, recorded observed over successive increments."

2. Objectives of time series.

Ans :

i) To evaluate past performance in respect of a particular variable.

ii) To make future forecasts in respect of the particular variable.

iii) To chart short term and long term strategies of the business in respect of the particular variable.

3. 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.

4. Least Square Method.

Ans :

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.

5. Index Numbers.

Ans :

Meaning

An Index, simply stated, is an indicator. It indicates the broad change in a given phenomenon. It is relative and indicates changes in the level of the given phenomenon in course of time, across geographical locations or in respect of any other characteristic. For example, the BSE Sensex (stands for Bombay Stock Exchange Sensitive Index) is an index of movement in stock prices. It tells us, very broadly, whether the prices of stocks of various companies have gone up or moved down on the Bombay Stock Exchange. Similarly, the Wholesale Price Index (WPI) gives us an indication of whether the general price level in the economy is going up or falling down. Similarly, Index numbers (or indices) can be constructed to measure any phenomenon, such as Industrial production, Number of road accidents, cost of Living or any other activity. Thus, Index Numbers are barometers measuring change in the level of a phenomenon.

Definitions

(i) According to Croton and Cowden Different Experts have defined Index Numbers in different words. Some of the definitions are stated as below.

(ii) According to Speiges "Index numbers are devices for measuring differences in the magnitude of a group of related variables."

6. Characteristics of Index Numbers.

Ans :

On the basis of the above definitions, the following points can be made:

1. Index Numbers are a measurement device
2. They do not measure or state the actual level attained by the phenomenon being studied. They measure the change in the phenomenon being studied.
3. The situations for which Index Numbers are used for comparison are not restricted in any manner. It can be a comparison of two time periods, two geographical locations, two groups of people or any other phenomenon.

7. Consumer Price Indexes

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. Report.

Ans :

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.

9. Reporting Writing.

Ans :

Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented and/or written. As a matter of fact even the most brilliant hypothesis, highly well designed and conducted research study, and the most striking generalizations and findings are of little value unless they are effectively communicated to others.

The purpose of research is not well served unless the findings are made known to others. Research results must invariably enter the general store of knowledge. All this explains the significance of writing research report. There are people who do not consider writing of report as an integral part of the research process.

10. Research Report.

Ans :

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.

Exercise Problems

1. The owner of Sun solar Heater is examining the number of solar homes strared in he region in each of the last 7 months :

Month	Aug.	Sept.	Oc.	Nov.	Dec.	Jan.	Feb.
No.of homes	15	15	26	27	33	41	51

Develop the linear equation and find the what will be the number of homes with solar heater in March..

2. Find the trend values by the method of he least squares for the following time series.

Year	1981	1982	1984	1985	1986	1989	1990	1992
Prod. ('000 units)	351	366	362	400	419	420	450	518

[Ans : Y = 406.84 + 21.23X: Y_{1995} = 579.91]

3. From the following data, calculate trend by 4 yearly average and find short-term oscillations :

Year	Production(Tons)
1984	5
1985	6
1986	7
1987	7
1988	6
1989	6
1990	9
1991	10
1992	9
1993	10
1994	11
1995	11

4. Fi a straight line trend to the following data by least squares method. Also find an estimate for the year 1997.

Year	1989	1990	1991	1992	1993	1994	1995	1996
No. of units ('000)	12	13	13	16	19	23	21	23

[Ans : Y = 17.5 + 893X, Y_{1197} = 25.54]

5. Fit a straight libe trend by the method of least square to the following data

Year	1990	1991	1992	1993	1994	1995	1996	1997
Sales(Thousand Rupees)	38	40	65	72	69	67	95	104

[Y = 68.75 + 4.404 X]

Choose the Correct Answer

1. Which of the following are the major characteristics of index numbers? [d]
   (a) It is expressed in percentages
   (b) It measures the net or relative changes in variables
   (c) It measures changes over a period of time
   (d) All of the above
2. The index number for base year is always ______. [c]
   (a) 1000 (b) 200
   (c) 100 (d) None of the above
3. The time period for which an index number is determined is known as ______. [c]
   (a) Base period (b) Normal period
   (c) Current period (d) None of the above
4. Index number is a type of ______. [c]
   (a) Dispersion (b) Correlation
   (c) Average (d) None of the above
5. Which of the following statements is/are correct about average prices if the price index is 110? [a]
   (a) The prices have increased by 10 per cent
   (b) The prices have increased by 110 per cent
   (c) The prices have decreased by 10 per cent
   (d) None of the above
6. Which of the following are limitations of using index numbers? [d]
   (a) The use of each index number is restricted to a specific object
   (b) It ignores the quality of commodities
   (c) It is useful only for short term comparison
   (d) All of the above
7. Which of the following is the variation within two or more variable studies by the index? [b]
   (a) Price index (b) Composite index
   (c) Simple index (d) None of the above
8. The weights used in a quantity index are ______. [c]
   (a) Quantity (b) Values
   (c) Price (d) None of the above
9. The report is always written in ______. [a]
   (a) Sequential manner (b) Irregular manner
   (c) Horizontal manner (d) Data biased manner
10. Report writing by the individual should be written in ______. [a]
    (a) First person (b) Last person
    (c) Both A, B (d) None

Fill in the Blanks

1. A ______ is a statistical data that are collected, observed or recorded at regular intervals of time.
2. ______ Trend is the basic tendency of a series to grow, decline, remain constant or fluctuate, over a long period of time.
3. ______ 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.
4. ______ are devices for measuring differences in the magnitude of a group of related variables.
5. ______ Method is similar to Laspeyre's method.
6. The circular test was proposed by ______.
7. CPIN Stands for ______.
8. The word 'report' is originated from the Latin word "report" which implies to ______.
9. ______ must be prepared and be given invariably in the report at the end.
10. ______ are notes placed at the bottom of a page.

ANSWERS

1. Time series
2. Secular
3. Least square
4. Index numbers
5. Paasche's
6. Weztergaard
7. Consumer Price Index Numbers
8. Carry back
9. Index
10. Footnotes

One Mark Answers

1. Seasonal Variations.

Ans :

Seasonal variations involve patterns of change within a year that tend to get repeated year after year.

2. Uses of trend analysis.

Ans :

i) The trend describes the basic growth tendency ignoring short term fluctuations.

ii) It describes the pattern of behavior which has characterized the series in the past.

3. Laspeyre's index method.

Ans :

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.

4. Time Reversal Test.

Ans :

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.

5. Factor Reversal Test.

Ans :

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.

χ_α^2 – Critical Values of the Chi-squared Distribution with v Degrees of Freedom

v	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.268
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.465
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.517
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.322
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.125
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.472	27.688	29.819	34.528
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.123
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.697
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.587	3Q. 191	30.995	33.409	35.718	40.790
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.820
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.315
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.797
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.620
26	29.246	30.434	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.052
27	30.319	31.528	32.912	36.741	40.113	43.194	44.140	46.963	49.645	55.476
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.993	56.893
29	32.461	33.71 1	35.139	39.087	42.557	45.772	46.693	49.588	52.336	58.302
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.703

Values of F for F Distribution at 1% Points
Degrees of freedom for numerator

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1.	4.052	5.000	5.403	5.625	5.764	5.859	5.928	5.982	6.023	6.056	6.106	6.157	6.209	6.235	6.261	6.287	6.313	6.339	6.366
2.	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3.	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
4.	21.2	28.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5.	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6.	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7.	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8.	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9.	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11.	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12.	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13.	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14.	8.68	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15.	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16.	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84.	2.75
17.	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18.	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19.	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20.	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21.	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22.	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23.	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24.	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25.	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30.	7.56	15.39	14.51	14.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40.	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60.	7.08	4.9	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120.	6.85	4.79	3.95	3.43	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.30
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.04	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

TABLE : CRITICAL VALUES OF STUDENT'S t-DISTRIBUTION

LEVEL OF SIGNIFICANCE FOR ONE-TAILED TEST

d.f.	.25	.10	.05	.025	.01	.005	.0005
	LEVEL OF S1GNIFICNCE FOR TWO-TAILED TEST						
v	.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

TABLE: CRITICAL VALUES OF THE VARIANCE RATIO *F*-DISTRIBUTION (RIGHT-TAIL AREAS) 2.5 PERCENT POINTS

υ_2 \ υ_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	674.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	976.71	984.87	993.10	997.25	1001.4	1005.6
2	38.51	39.00	39.17	39.25	39.30	39.33	39.33	39.37	39.39	39.40	39.42	39.43	39.45	39.46	39.47	39.47
3	17.44	16.04	15.44	15.01	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.91	8.84	8.75	8.66	8.56	8.51	8.46	8.41
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91
13	6.41	4.97	4.35	3.99	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25
22	5.79	4.38	3.78	3.44	3.22	3.05	2.94	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.95	1.82	1.76	1.69	1.61
∞	5.00	3.67	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.85	1.71	1.64	1.57	1.48

υ_1 : d.f. for the numerator ; υ_2 : d.f. for the denominator.

Internal Assessment (Mid Examinations)

In CIE, for theory subjects, during a semester, there shall be two mid-term examinations. Each MidTerm examination consists of two parts i) **Part – A** for 10 marks, ii) **Part – B** for 15 marks with a total duration of 2 hours as follows:

1. Mid-Term Examination for 25 marks:

   (a) Part - A: Objective/quiz paper/Short Note for 10 marks.

   (b) Part - B: Descriptive paper for 15 marks.

   Student shall have to earn 40%, i.e. 10 marks out of 25 marks from average of two mid-term examinations (I Mid-Term & II Mid-Term).

   The remaining 15 marks of Continuous Internal Assessment (out of 40) are distributed as:

2. Assignment for 5 marks. (Average of 2 Assignments each for 5 marks)

3. PPT/Poster Presentation/ Case Study/Video presentation/Survey/Field Study/Group discussion /Role Play on a topic in the concerned subject for $5+5 = 10$ marks before II MidTerm Examination.

   - The objective/quiz paper is set with multiple choice, fill-in the blanks, match the following type of questions and short notes for a total of 10 marks. The descriptive paper shall contain 5 full questions out of which, the student has to answer 3 questions, each carrying 5 marks. The student has to get minimum of 40% (on 25 marks allocated for Mid-Term examinations) on average of two Mid-Term examinations.

   - While the first mid-term examination shall be conducted on 50% of the syllabus, the second mid-term examination shall be conducted on the remaining 50% of the syllabus.

   - Five (5) marks are allocated for assignments (as specified by the subject teacher concerned). The first assignment should be submitted before the conduct of the first mid-term examination, and the second assignment should be submitted before the conduct of the second mid-term examination.

   - The average of the two assignments shall be taken as the final marks for assignment (for 5 marks). PPT/Poster Presentation/ Case Study/ Video presentation/ Survey/ Field Study/ Group discussion / Role Play on a topic in the concerned subject for $5+5 = 10$ marks before II Mid-Term Examination.

UNIT - I

Part - A

Multiple Choice Questions

1. Who was the author of the book named "Methods in Social Research"? [c]

   (a) Kerlinger (b) CR Kothari

   (c) Goode and Hatt (d) Wilkinson

2. A research intends to explore the result of possible factors for the organization of effective mid-day meal interventions. Which research method will be most appropriate for this study? [c]
   (a) Descriptive survey method (b) Historical method
   (c) Ex-post facto method (d) Experimental method
3. The main objective of ______ study's to acquire knowledge [b]
   (a) Exploratory (b) Descriptive
   (c) Diagnostic (d) Descriptive and Diagnostic

Fill in the Blanks

4. ______ research aims at discovery of new facts and verifications of old ones. **(Research)**
5. ______ research involves the develop- ment of new theories, abstract ideas, and generalized principles. **(Conceptual)**
6. ______ scales have an absolute or true zero of measurement. **(Ratio)**

Short Notes

7. Define Research. **(Unit-I, SQA - 1)**
8. Analytical Research. **(Unit-I, SQA - 7)**
9. Exploratory Research. **(Unit-I, SQA - 9)**
10. Variable. **(Unit-I, SQA - 10)**

Part - B

1. What are the features of good research. **(Unit-I, Q.No. 2)**
2. What are the unique challenges in conduct of social science research ? **(Unit-I, Q.No. 9)**
3. Define the term variable. Explain various types of variables. **(Unit-I, Q.No. 16)**
4. Define measurement. What are the characteristics of measurement? **(Unit-I, Q.No. 18)**
5. Differentiate among nominal, ordinal, interval and ratio scales on any five points. **(Unit-I, Q.No. 23)**

UNIT - II

Part - A

Multiple Choice Questions

1. Survey is a ______ Study. [b]
   (a) Descriptive (b) Fact finding
   (c) Analytical (d) Systematic
2. What are the core elements of a Research Process? [d]
   (a) Introduction; Data Collection; Data Analysis; Conclusions and Recommendations
   (b) Executive Summary; Literature Review; Data Gathered; Conclusions; Bibliography
   (c) Research Plan; Research Data; Analysis; References
   (d) Introduction; Literature Review; Research Methodology; Results; Discussions and Conclusions

3. Questionnaire is filled by ______. [a]
   (a) Respondent (b) Everybody
   (c) Enumerator (d) None of the above

Fill in the Blanks

4. A ________ is a specific statement in the general area of investigation. **(Research problem)**
5. ________ research is characterized by a high degree of flexibility. **(Exploratory)**
6. ________ designs are used in experiments where the effect of varying more than one factor are to be determined. **(Factorial)**

Short Notes

7. What is a research design? **(Unit-II, SQA - 2)**
8. Exploratory Research Design. **(Unit-II, SQA - 4)**
9. Population. **(Unit-II, SQA - 6)**
10. What is Sampling? **(Unit-II, SQA - 5)**

Part - B

1. Briefly explain Research Problems in the fields of marketing, production, finance and personnel management. **(Unit-II, Q.No. 4)**
2. What are the features of Good Research Design? **(Unit-II, Q.No. 11)**
3. Briefly explain random and non-random sampling. **(Unit-II, Q.No. 15)**
4. What are the methods for collecting secondary data? **(Unit-II, Q.No. 24)**
5. What are the various sources of errors in measurement? **(Unit-II, Q.No. 39)**

UNIT - III

Part - A

Multiple Choice Questions

1. When the quantitative and qualitative data are arranged according to a single feature, what is the tabulation known as? [a]
   (a) One-way (b) Bivariate
   (c) Manifold division (d) Dichotomy
2. What is the table where the variables are subdivided with interrelated features known as? [d]
   (a) Order level table (b) Sub-parts of a table
   (c) One-way table (d) Two-way table
3. The runs scored by a batsman in 5 ODIs are 31,97,112, 63, and 12. The standard deviation is [c]
   (a) 24.79 (b) 23.79
   (c) 25.79 (c) 26.79

Fill in the Blanks

4. The ________ selected must satisfy all the values to be plotted on the graph. **(Scale)**

5. A ________ Diagram is a type of graph that displays data in a circular graph. **(Pie)**

6. ________ is the simplest method of studying dispersion. **(Range)**

Short Notes

7. Diagrams. **(Unit-III, SQA - 4)**

8. Differentiate between diagrams and graphs. **(Unit-III, SQA - 6)**

9. Define mode. **(Unit-III, SQA - 9)**

10. Define Dispersion. **(Unit-III, SQA - 10)**

Part - B

1. Explain in detail the differnet parts of tables. **(Unit-III, Q.No. 5)**

2. For the following data, draw Histogram, Frequency Curve and frequency Polygon.

Salary (Rs. in '000)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Employees	15	20	30	50	40	20	10

(Unit-III, Prob. 8)

3. Explain the functions of statistics. **(Unit-III, Q.No. 24)**

4. Explain the properties and applications of t-distribution. **(Unit-III, Q.No. 65)**

5. Two different methods of training the students in an institute gave the following results of marks.

Method A	17	19	18	15	21	19	16	14	10	15
Method B	20	16	21	17	19	16	16	15		

Test whether the two methods are significantly different at $\alpha = 0.05$.

(Unit-III, Prob. 40)

UNIT - IV

Part - A

Multiple Choice Questions

1. Analysis of variance is a statistical method of comparing the of several populations. [a]

   (a) Means (b) Variances

   (c) Standard Deviations (d) None of The Above

2. When conducting an ANOVA, FDATAwill always fall within what range? [a]

   (a) between 0 and infinity (b) between 0 and 1

   (c) between negative infinity and infinity (d) between 1 and infinity

3. Which of the following statements is true for correlation analysis? [c]
   (a) It is a bivariate analysis (b) It is a multivariate analysis
   (c) It is a univariate analysis (d) Both a and c

Fill in the Blanks

4. An ________ means a quality or characteristic. **(Attribute)**
5. The term 'regression' was first used by ________ in 1877 **(Sir Francis Galton)**
6. ________ helps in establishing a functional relationship between two or more variables. **(Regression analysis)**

Short Notes

7. Chi-Square Distribution. **(Unit-IV, SQA - 2)**
8. Spearman's Rank Correlation. **(Unit-IV, SQA - 7)**
9. Demerits. **(Unit-IV, SQA - 6)**
10. Regression. **(Unit-IV, SQA - 8)**

Part - B

1. What is an ANOVA? State the assumptions and applications of an ANOVA. **(Unit-IV, Q.No. 1)**
2. A farmer applied three types of fertilizers on 4 separate plots. The figure on yield per acre are tabulated below:

Fertilizers Plo 2 ts	Yield				
	A	B	C	D	Total
Nitrogen	6	4	8	6	24
Potash	7	6	6	9	28
Phosphates	8	5	10	9	32

   Find out if the plots are materials different in fertility, as also, if the three fertilizers make materials difference in yields. **(Unit-IV, Prob. 1)**
3. Explain briefly about two way ANOVA with and without interation. **(Unit-IV, Q.No. 3)**
4. What are the different types of correlations? **(Unit-IV, Q.No. 10)**
5. Define multiple regression analysis. State its objectives and assumptions. **(Unit-IV, Q.No. 24)**

UNIT - V

Part - A

Multiple Choice Questions

1. Which of the following are the major characteristics of index numbers? [d]
   (a) It is expressed in percentages
   (b) It measures the net or relative changes in variables
   (c) It measures changes over a period of time
   (d) All of the above

2. Which of the following is the variation within two or more variable studies by the index? [b]

(a) Price index (b) Composite index

(c) Simple index (d) None of the above

3. Which of the following are limitations of using index numbers? [d]

(a) The use of each index number is restricted to a specific object

(b) It ignores the quality of commodities

(c) It is useful only for short term comparison

(d) All of the above

Fill in the Blanks

4. ______ Trend is the basic tendency of a series to grow, decline, remain constant or fluctuate, over a long period of time. **(Secular)**

5. ______ are devices for measuring differences in the magnitude of a group of related variables. **(Index numbers)**

6. CPIN Stands for ______. **(Consumer Price Index Numbers)**

Short Notes

7. Define time series. **(Unit-V, SQA - 1)**

8. Characteristics of Index Numbers. **(Unit-V, SQA - 6)**

9. Report. **(Unit-V, SQA - 8)**

10. Research Report. **(Unit-V, SQA - 10)**

Part - B

1. What are the Components of Time Series ? **(Unit-V, Q.No. 4)**

2. Describe briefly applications of index numbers in business and industry. **(Unit-V, Q.No. 16)**

3. 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. **(Unit-V, Prob. 10)**

4. Define Report, Report Writing and Research Report. State the importance of Report Writing. **(Unit-V, Q.No. 27)**

5. What is Documentation. Explain the Guides for Effective Documentation. **(Unit-V, Q.No. 34)**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

M.B.A I - Year I - Semester Examination

R22

MODEL PAPER - I

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 3 Hours] **[Max. Marks : 60**

Note : This question paper contains two parts **A** and **B**.

Part A is compulsory which carries 10 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 (10 × 1 = 10 Marks)

ANSWERS

1. (a) Define Research. **(Unit - I, SQA -1)**
   (b) Who are the stakeholders in research ? Who are the participants in research? **(Unit - I, SQA - 3)**
   (c) What is a research design? **(Unit - II, SQA - 2)**
   (d) Quota Sampling. **(Unit - II, SQA -7)**
   (e) Diagrammatic Representation of Data **(Unit - III, SQA - 3)**
   (f) Testing for One and Two Means **(Unit - III, SQA - 13)**
   (g) What is an ANOVA **(Unit - IV, SQA -1)**
   (h) Define Correlation. **(Unit - IV, SQA - 4)**
   (i) Report. **(Unit - V, SQA - 8)**
   (j) Objectives of time series. **(Unit - V, SQA - 2)**

PART - B (5 × 10 = 50 Marks)

2. (a) What are the unique challenges in conduct of social science research ? **(Unit - I, Q.No.9)**

   OR

   (b) Briefly explain the step-by-step process of research. **(Unit - I, Q.No.13)**

3. (a) Briefly explain Research Problems in the fields of marketing, production, finance and personnel management. **(Unit - II, Q.No.4)**

   OR

   (b) Briefly explain random and non-random sampling. **(Unit - II, Q.No. 15)**

4. (a) 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 ssumption 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. **(Unit - III, Prob.No.37)**

OR

(b) Explain the functions of statistics. **(Unit - III, Q.No.24)**

5. (a) 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. **(Unit - IV, Prob.No.3)**

OR

(b) Explain briefly about Least Square Fit of Linear Regression. **(Unit - IV, Q.No. 23)**

6. (a) Fit a straight line by the Least Square Method and tabulate the trend values for the above data.

Year	2011	2012	2013	2014	2015	2016	2017
Production (in tons)	77	88	94	85	91	98	90

(Unit - V, Prob. No.7)

OR

(b) Discuss various types of Research Reports. **(Unit - V, Q.No.28)**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

M.B.A I - Year I - Semester Examination **R22**

MODEL PAPER - II

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 3 Hours] **[Max. Marks : 60**

Note : This question paper contains two parts **A** and **B**.

Part A is compulsory which carries 10 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 (10 × 1 = 10 Marks)

ANSWERS

1. (a) Features of good research. **(Unit - I, SQA - 2)**

(b) Analytical Research **(Unit - I, SQA -7)**

(c) What are Research Problems? **(Unit - II, SQA - 1)**

(d) Information. **(Unit - II, SQA - 9)**

(e) Diagrams **(Unit - III, SQA - 4)**

(f) Define Dispersion. **(Unit - III, SQA -10)**

(g) Chi-Square Distribution. **(Unit - IV, SQA - 2)**

(h) Spearman's Rank Correlation. **(Unit - IV, SQA - 7)**

(i) Reporting Writing. **(Unit - V, SQA - 9)**

(j) Define time series. **(Unit - V, SQA -1)**

PART - B (5 × 10 = 50 Marks)

2. (a) Take an Example of doing market research before Launching a new product. **(Unit - I, Q.No.14)**

OR

(b) What are the characteristics of good measurement tool? **(Unit - I, Q.No.18)**

3. (a) What are the Applications of Sampling? **(Unit - II, Q.No.18)**

OR

(b) What are the methods for collecting secondary data? **(Unit - II, Q.No.24)**

4. (a) 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.No.6)

OR

(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). **(Unit - III, Prob.No.38)**

5. (a) 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.No. 4)**

OR

(b) Find if there is any significant correlation between the heights and weights given below:

Height in inches	57	59	62	63	64	65	55	58	57
Weight in lbs	113	117	126	126	130	129	111	116	112

(Unit - IV, Prob.No.11)

6. (a) The following table gives the annual sales (in Rs.'000) of a commodity :

Year	Sales
1990	710
1991	705
1992	680
1993	687
1994	757
1995	629
1996	644
1997	783
1998	781
1999	805
2000	805

Determine the trend by calculating the 5-yearly moving average. **(Unit - V, Prob.No. 3)**

OR

(b) What is Documentation. Explain the Guides for Effective Documentation. **(Unit - V, Q.No.34)**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

M.B.A I - Year I - Semester Examination **R22**

MODEL PAPER - III

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 3 Hours] **[Max. Marks : 60**

Note : This question paper contains two parts **A** and **B**.

Part A is compulsory which carries 10 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 (10 × 1 = 10 Marks)

ANSWERS

1. (a) Who are the stakeholders in research ? Who are the participants in research? **(Unit - I, SQA - 3)**
   (b) Variable **(Unit - I, SQA - 10)**
   (c) Purpose of Research Design. **(Unit - II, SQA - 3)**
   (d) What is Sampling? **(Unit - II, SQA - 5)**
   (e) Define Classification. **(Unit - III, SQA - 1)**
   (f) Graphical representation of data **(Unit - III, SQA - 5)**
   (g) Use of Chi-Square Test. **(Unit - IV, SQA - 3)**
   (h) Define multiple regression analysis. **(Unit - IV, SQA - 10)**
   (i) Consumer Price Indexes **(Unit - V, SQA - 7)**
   (j) Least Square Method. **(Unit - V, SQA - 4)**

PART - B (5 × 10 = 50 Marks)

2. What are the applications and disadvantages of nominal, ordinal, interval and ratio scales. **(Unit - I, Q.No.22)**

OR

   (b) What are the Ethical issues concerning research Participants? **(Unit - I, Q.No.24)**

3. (a) Define Questionnaire. Explain different types of Questionnaire. **(Unit - II, Q.No. 28)**

OR

   (b) What are the Guidelines / Precautions for Questionnaire ? **(Unit - II, Q.No. 31)**

4. (a) What do you mean by graphical repre-sentation of data ? State the rules for graphing. What are the advantages and disadvantages of graphs ? **(Unit - III, Q.No.13)**

OR

(b) Calculate Standard Deviation and Coefficient of Variation from the following data:

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
No. of Students	5	7	14	28	12	9	6	2

(Unit - III, Q.No.33)

5. (a) 200 digits were chosen at random from a set of tables. The frequencies of the digits are shown below :

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	18	19	23	21	16	25	22	20	21	15

Use the chi-square test to assess the correctness of the hypothesis that the digits were distributed in equal number in the tables from which these were chosen. **(Unit - IV, Prob.No. 8)**

OR

(b) What is Discriminant Analysis? Explain the objectives of discriminant analysis. **(Unit - IV, Q.No. 28)**

6. (a) Construct index numbers of price from the following data by applying :

(i) Laspeyres method

(ii) Paasche method

(iii) Fisher's ideal method, and

(iv) Marshall-Edgeworth method

	2010		**2011**	
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

(Unit - V, Prob.No. 12)

OR

(b) Explain the process of Report Preparation and Presentation. **(Unit - V, Q.No.29)**

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.21)**

   (c) Differentiate between univariate and multivariate data. **(Unit-III, Q.No.8)**

   (d) When is a t-test used? What are its different types? **(Unit-III, Q.No.12,13,14,15)**

   (e) What are some problems that are encountered while constructing index numbers? **(Unit-V, Q.No.23)**

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.1,3,11)**

OR

3. Describe the research process in detail. Take an example of doing market research before launching a new product. **(Unit-I, Q.No.13,14)**

4. What are the features of a good research design? **(Unit-II, Q.No.11)**

OR

5. What is a research design? Discuss the different types of common research designs. **(Unit-II, Q.No.5, 7)**

6. Define tabulation. Explain in delail the different parts of a table. **(Unit-III, Q.No.4, 5)**

OR

7. What is a dependent Sample or repeated measures t-test? Explain its use by giving a suitable example. **(Unit-III, Q.No.64, 68)**

8. What is ANOVA? How is an ANOVA table setup? **(Unit-IV, Q.No.1, 2)**

OR

9. Define Correlation. Explain the significance of Correlation. **(Unit-IV, Q.No.9)**

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.33)**

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. **(Unit-V, Prob.4)**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

MBA I-Semester Examinations **R19**

October / November - 2020

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.18,22)**

   (b) What are the unique challenges in conduct of social science research ? **(Unit-I, Q.No.9)**

2. (a) Who are the stakeholders in research ? Who are the participants in research? **(Unit-I, Q.No.4)**

   (b) Distinguish between the terms 'reliability' and 'validity'. **(Unit-II, Q.No.38)**

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.8,37)**

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.9)**

   (b) Explain how 'ogives' are drawn for any frequency distribution. **(Unit-III, Q.No.10)**

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, Prob.39)**

   (b) What is

   (i) pie-chart **(Unit-III, Q.No.19)**

   (ii) histogram ? **(Unit-III, Q.No.14 (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. **(Unit-IV, Prob.3)**

7. (a) What care must be taken in designing capitalization in report preparation? **(Unit-V, Q.No.35)**

   (b) What is correlation? **(Unit-IV, Q.No.9)**

   (c) What are the components of 'time series'" **(Unit-V, Q.No.4)**

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. **(Unit-V, Prob.10)**

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 ? **(Unit-I, Q.No.9)**

   (b) What are the levels of measurement ? **(Unit-I, Q.No.22)**

2. (a) What are the characteristics of good measurement tool ? **(Unit-I, Q.No.18)**

   (b) What is meant by exploratory research? What are the purposes for which it is conducted? **(Unit-I, Q.No.12)**

3. (a) What are the sources of research problem ? **(Unit-II, Q.No.3)**

   (b) Explain "Before-after experimental design with control group". **(Unit-II, Q.No.10)**

4. (a) Why is graphic representation of data desirable ? **(Unit-III, Q.No.13)**

   (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). **(Unit-III, Prob.38)**

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

(Unit-IV, Prob.10)

6. (a) What is cluster analysis ? What are its applications ? **(Out of Syllabus)**

   (b) Explain clearly the differences between simple linear and multiple linear regressions. **(Unit-IV, Q.No.27)**

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

(Unit-V, Prob.9)

8. (a) What are the purpose of footnotes? What is the kind of footnotes one finds in research. **(Unit-V, Q.No.33)**

(b) What is an index number? Describe briefly in application in business and industry. **(Unit-V, Q.No.13,16)**

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 ? **(Unit-I, Q.No.9)**

   (b) What are the ethical issues concerning research participants ? **(Unit-I, Q.No.24)**

2. (a) What is a diagnostic study ? What is the purpose of it ? **(Unit-I, Q.No.26)**

   (b) Distinguish between questionnaire and schedule for collection of data. **(Unit-II, Q.No.34)**

3. What is secondary source of data ? Critically evaluate its use for research. What are the limitations? **(Unit-II, Q.No.24,25)**

4. (a) What is multivariate table ? Illustrate your answer with a specimen copy of table containing fictitious data. **(Unit-III, Q.No.7)**

   (b) Under what conditions can we apply 't' test for test of hypothesis ? **(Unit-III, Q.No.64,65)**

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. **(Unit-III, Prob.37)**

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 ? **(Unit-III, Prob.41)**

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. **(Unit-IV, Prob.2)**

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. **(Unit-V, Prob.8)**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

MBA I - Semester Examinations

R19

May - 2022

RESEARCH METHODOLOGY AND STATISTICAL ANALYSIS

Time : 3 Hours] **[Max. Marks : 75**

Answer any Five questions
(All questions carry equal marks)

ANSWERS

1. (a) Briefly explain the step by step process of Research. **(Unit-I, Q.No.13)**

   (b) What are the four types of Measurement of Variables? What are the applications and disadvantages of Nominal, Ordinal, Interval and Ratio Scales? **(Unit-I, Q.No.22)**

2. (a) What is Research Process? Briefly describe the steps involved in Business Research Process. **(Unit-I, Q.No.13)**

   (b) Differentiate among Nominal, Ordinal, Interval and Ratio Scales on any five points. **(Unit-I, Q.No.23)**

3. (a) What are Research Problems? Briefly explain Research Problems in the fields of Marketing Production, Finance and Personnel Management. **(Unit-II, Q.No.1,4)**

   (b) What are the Methods for collecting Secondary Data? What is the significance of Secondary Data? **(Unit-II, Q.No.24,25)**

4. (a) What is Bias in Survey? What are the reasons for biases? Explain the the same briefly. **(Unit-II, Q.No.17)**

   (b) What are the guidelines for Questionnaire? Briefly Explain Random and Non-Random sampling with suitable examples. **(Unit-II, Q.No.31)**

5. (a) What is Frequency Distribution? Prepare the Cumulative Frequency Table for the following data : **(Unit-III, Q.No.9, Prob.1)**

Marks	10-29	30-49	50-69	70-89	90-100
No. of Students	5	25	45	20	5

   (b) For the following data, prepare a Bar Diagram : **(Unit-III, Prob.9)**

Year	2014	2015	2016	2017	2018	2019	2020	2021
Property Tax (Rs. Crore)	2000	2500	3000	3600	4700	5800	6000	6500

6. (a) For the following data, draw Histogram, Frequency Curve and frequency Polygon. **(Unit-III, Prob.8)**

Salary (Rs. in '000)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Employees	15	20	30	50	40	20	10

(b) For the given data below on the monthly income of 600 families of a rich locality in Hyderabad City, draw :

(i) Less than Ogive and

(ii) More than Ogive Curves and point out the Median. **(Unit-III, Prob.2)**

Monthly Income (Rs. Crore)	No. of Families
Below 10	50
10 - 20	100
20 - 30	200
30 - 40	150
40 - 50	50
50 - 60	30
60 - Above	20

7. (a) What is Multiple Regression Analysis? What are its Objectives? Explain its Formula. What is its statistical Significance? **(Unit-IV, Q.No.24,25)**

(b) What is Spearman's Rank Correlation? Solve the following problem on Spearman's Rank Correlation. **(Unit-IV, Prob.17)**

Candidate	Judge P	Judge Q
1	20	24
2	23	24
3	25	22
4	24	26
5	26	30

8. (a) Give a brief introduction of Index Numbers What are their uses? Construct Laspayre and Paasche Index Numbers for the following data:

Commodity	2020 Price	2020 Quantity	2021 Price	2021 Quantity
P	3	10	5	7
Q	6	12	7	6
R	5	16	6	10
S	3	20	3	12

(Unit-V, Q.No.13,15, Prob.14)

(b) What is the concept and importance of Research Report-writing? **(Unit-V, Q.No.27)**