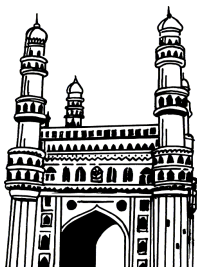
Rahul's ✔

Topper's Voice



M.Com.

II Year III Semester

Latest 2024 Edition

RESEARCH METHODOLOGY AND QUANTITATIVE TECHNIQUES

- ☞ **Study Manual**
- ☞ **FAQ's and Important Questions**
- ☞ **Exercise Problems**
- ☞ **Short Questions & Answers**
- ☞ **Choose the Correct Answers**
- ☞ **Fill in the blanks**
- ☞ **One Mark Answers**
- ☞ **Solved Model Papers**

Price ₹ 199/-

- by -

WELL EXPERIENCED LECTURER

***Rahul Publications*™**

Hyderabad. Cell : 9391018098, 9505799122.

M.Com.

II Year III Semester

RESEARCH METHODOLOGY AND QUANTITATIVE TECHNIQUES

Inspite of many efforts taken to present this book without errors, some errors might have crept in. Therefore we do not take any legal responsibility for such errors and omissions. However, if they are brought to our notice, they will be corrected in the next edition.

Price : ` 199/-

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 QUANTITATIVE TECHNIQUES

STUDY MANUAL

SOLVED MODEL PAPERS

SYLLABUS

UNIT - I

Introduction, Research: Meaning, Purpose, Characteristics and Types - Process of Research: Formulation of objectives - Research plan and its components - Methods of Research: Survey, Observation, Case study, Experimental, Empirical and Comparative Methods - Difficulties in Business Research (Theory only)

Collection, Presentation & Analysis of data: Sources of Data: Primary and Secondary Sources - Methods of collecting Primary Data - Measurement and Scaling -Designing Questionnaires/Schedules. Census vs. Sampling - Methods of Sampling Random and Non-Random Sampling methods - Techniques. (Theory only)

UNIT - II

Interpretation: Introduction - Essentials for Interpretation, Precautions in interpretation - Conclusions and generalization - Methods of generalization. Statistical fallacies: bias, inconsistency in definitions, inappropriate comparisons, faulty generalizations, drawing wrong inferences, misuse of statistical tools, failure to comprehend the data. (including small cases).

Report Writing: Meaning and types of reports - Stages in preparation of Report - Characteristics of a good report - Structure of the report'-Documentation: Footnotes and Bibliography - Checklist for the report.

UNIT - III

Statistical Estimation: Concepts of Population, Sample and Sampling Distribution- Parameters and Statistics - Central limit theorem - Concept of Standard Error - Confidential Limits - Estimation of Population Parameters - Properties of a Good Estimator - Point and Interval Estimation.

Hypothesis Testing: Formulation of Hypotheses: Types of Hypotheses - Methods of testing Hypotheses - Type I and Type II errors - One tail and two tail tests (Theory only). Sampling of Attributes: Estimation and testing of Number and Proportions of Success - Difference between two proportions (including problems).

Sampling of Variables: Large Samples: Difference between large and small samples - Estimating population mean - Testing: Significance of Mean - Significance of the difference between means of two samples - Significance of the difference between the standard deviations of two samples. (Including problems) Small Samples: 't' test - Fixing fiducial limits to population mean – Testing: Significance of the mean - Significance of the difference between two independent means – Significance of the difference between two dependent means (including problems).

UNIT - IV

Analysis of Variance: F-test: Meaning and Applications - ANOVA: Assumptions - Procedure - One way and two-way analysis of variance (including Problems).

Association of Attributes & Chi-Square Test: Association of Attributes: Meaning- Distinction between Correlation and Association - Methods of studying Association - Interpretation of results.

Chi Square Test: Definition - Conditions for applying Chi square test, Yates s correction– Uses and limitations of Chi square test – Chi square test for testing the independence of Attribute – Chi square test for goodness of fit (including problems).

UNIT - V

Statistical Decision Theory: Nature of Decision - State of Nature – Pay off Tables - Expected Pay off - Expected Opportunity Loss – Value of Perfect Information – Types of Decision Situation – Choice of Decision Criteria – Decision Tree Analysis – Decision Making under Uncertainty (including simple problems).;

Game Theory: Characteristics of Game Theory – Two Persons Zero Sum Game - Maximum and Minimax Strategies – Saddle Point – Dominating Strategy – Mixed Strategy - Limitations of Game Theory (including simple problems with Analytical Formulae and Graphical Methods).

Linear Programming: Linear Programming: Meaning - Requirements for application - Assumptions - Advantages - Application of LP - Formulation of LP problems (including simple problems). Graphical Solutions of LP problems with two variables only (including simple problems).

Contents

UNIT - I

UNIT - II

UNIT - V

Frequently Asked & Important Questions

UNIT - I

1. Define the term Research. Explain the purpose of Research.

Ans : **(Imp.)**

Refer Unit-I, Page No. 1, Q.No. 1.

---

2. Briefly describe the different steps involved in a Research Process.

Ans : **(Imp.)**

Refer Unit-I, Page No. 4, Q.No. 4.

---

3. Describe the various methods of Research.

Ans : **(Imp.)**

Refer Unit-I, Page No. 8, Q.No. 7.

---

4. What are the difficulties involved in business research?

Ans : **(Imp.)**

Refer Unit-I, Page No. 9, Q.No. 8.

---

5. Critically examine the different methods of collecting primary data.

Ans : **(Dec.-18, Dec.-16, Dec.-13, Imp.)**

Refer Unit-I, Page No. 10, Q.No. 10.

---

6. Explain the various types of measurement scales with example.

Ans : **(Imp.)**

Refer Unit-I, Page No. 14, Q.No. 15.

---

7. Distinguish between nominal, ordinal, Interval and Ratio Scale.

Ans : **(Imp.)**

Refer Unit-I, Page No. 15, Q.No. 16.

---

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

Ans : **(Imp.)**

Refer Unit-I, Page No. 18, Q.No. 20.

---

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

Ans : **(Imp.)**

Refer Unit-I, Page No. 19, Q.No. 21.

10. Distinguish between questionnaire and schedule.

Ans : **(Imp.)**

Refer Unit-I, Page No. 21, Q.No. 23.

11. Explain briefly the various random sampling methods.

Ans : **(Dec.-16, Imp)**

Refer Unit-I, Page No. 23, Q.No. 26.

UNIT - II

1. Why is Interpretation considered a basic component of research process?

Ans : **(Dec.-18, Dec.-16, Dec.-14, Imp.)**

Refer Unit-II, Page No. 33, Q.No. 1.

2. Explain the various methods of genera- lization.

Ans : **(Dec.-15, Imp.)**

Refer Unit-II, Page No. 34, Q.No. 4.

3. State the different statistical fallacies that may arise while interpreting the data. Explain each one of them with relevant examples.

Ans : **(Dec.-18, Dec.-16, June-16, Dec.-14, Dec.-13, Imp.)**

Refer Unit-II, Page No. 36, Q.No. 5.

4. What is a Report and Report writing. State its objectives.

Ans : **(Dec.-14, Imp.)**

Refer Unit-II, Page No. 37, Q.No. 6.

5. Explain the various stages in preparation of report.

Ans : **(June-16, Dec.-14, Imp.)**

Refer Unit-II, Page No. 39, Q.No. 8.

6. Out line the structures of a Research Report.

Ans : **(Dec.-18, Dec.-13, Imp.)**

Refer Unit-II, Page No. 41, Q.No. 10.

7. Define and discuss footnotes in Research Report. Explain the purposes of footnotes.

Ans : **(Dec.-18, Dec.-16, Imp.)**

Refer Unit-II, Page No. 43, Q.No. 11.

8. Explain the Checklist for the Research Report.

Ans : **(Imp.)**

Refer Unit-II, Page No. 44, Q.No. 13.

UNIT - III

1. Describe briefly about the concept of standard error and confidential limits.

Ans : **(Imp.)**

Refer Unit-III, Page No. 54, Q.No. 3.

2. State the properties of a Good Estimator.

Ans : **(Imp.)**

Refer Unit-III, Page No. 56, Q.No. 5.

3. Explain various methods of Testing Hypothesis.

Ans : **(Imp.)**

Refer Unit-III, Page No. 59, Q.No. 8.

4. Explain briefly about One Tailed and Two Tailed Tests.

Ans : **(Imp.)**

Refer Unit-III, Page No. 60, Q.No. 10.

5. What do you understand by Sampling of Attributes?

Ans : **(Imp.)**

Refer Unit-III, Page No. 63, Q.No. 12.

6. 500 apples are taken at random from a large basket and 50 are found to be bad. Estimate the proportion of bad apples in the basket and assign limits within which the percentage most probably lies.

Sol : **(Imp.)**

Refer Unit-III, Page No. 65, Prob. 3.

7. Define the term large sample. What are the assumptions of large samples.

Ans : **(Imp.)**

Refer Unit-III, Page No. 68, Q.No. 17.

8. Explain briefly about hypothesis concerning two mean large samples.

Ans : **(Imp.)**

Refer Unit-III, Page No. 73, Q.No. 21.

9. Define 't' distribution. Explain the properties and applications of 't' distribution.

Ans : **(Imp.)**

Refer Unit-III, Page No. 79, Q.No. 24.

10. The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data

Item :	1	2	3	4	5	6	7	8	9	10
Life in '000 hour :	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average life time of bullbs is 4,000 hours.

Sol : **(Imp.)**

Refer Unit-III, Page No. 82, Prob. 14.

11. Eleven sales executive trainees are assigned selling jobs right after their recruitment. After a fortnight they are withdrawn from their field duties and given a month's training for executives sales. Sales executed by them in thousands of rupees before and after the training in the same period are listed below.

Sales (000 `) (Before training)	23	20	19	21	18	20	18	17	23	16	19
Sales (000 `) (After training)	24	19	21	18	20	22	20	20	23	20	27

Do these data indicate that the training has contributed to their performance?

Sol : **(Imp.)**

Refer Unit-III, Page No. 87, Prob. 17.

UNIT - IV

1. Define is ANOVA? What are the assumptions and applications of ANOVA.

Ans : **(Imp.)**

Refer Unit-IV, Page No. 100, Q.No. 3.

2. Enlist the techniques of analysis ot variance. Explain about briefly one way ANOVA.

Ans : **(Imp.)**

Refer Unit-IV, Page No. 100, Q.No. 4.

3. 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 : **(Imp.)**

Refer Unit-IV, Page No. 106, Prob. 3.

4. Define correlation. Explain the significance of correlation.

Ans : **(Imp.)**

Refer Unit-IV, Page No. 111, Q.No. 7.

5. Explain Yule's coefficient of association.

Ans : **(Imp.)**

Refer Unit-IV, Page No. 117, Q.No. 13.

6. What is Chi-Square (χ^2)? Explain the applications test statistic and assumptions.

Ans : **(Imp.)**

Refer Unit-IV, Page No. 121, Q.No. 16.

7. Explain briefly about Yates's correction.

Ans : **(Imp.)**

Refer Unit-IV, Page No. 123, Q.No. 18.

8. Chi Square Test is a test of Independence, Homogeneity and Goodness of fit. Discuss briefly.

Ans : **(Imp.)**

Refer Unit-IV, Page No. 124, Q.No. 19.

9. Explain briefly about Chi square test for testing the independence of Attribute.

Ans : **(Imp.)**

Refer Unit-IV, Page No. 125, Q.No. 20.

10. Four machines A, B, C, D are used to manufacture parts which are classified as first, second and third grade. Test whether the quality of the product is independent of the machines. The data as under,

Grade	A	B	C	D
First	620	750	400	530
Second	130	200	140	130
Third	50	50	60	40

Sol : **(Imp.)**

Refer Unit-IV, Page No. 127, Prob. 14.

11. The shopkeeper feels that daily demand of a product follows uniform distribution. The observed frequencies of demand values are as follows:

Observed frequencies of daily demand

Demand (Units)	Observed Frequency (O_1)	Demand (Units)	Observed Frequency (O_1)
20	13	26	12
21	10	27	10
22	7	28	14
23	10	29	9
24	6		
25	9		

Check whether the given data follow uniform distribution at a significance level of 0.05.

Sol : **(Imp.)**

Refer Unit-IV, Page No. 133, Prob. 16.

UNIT - V

1. Elaborate the various steps involved in decision making.

Ans : **(Imp.)**

Refer Unit-V, Page No. 146, Q.No. 3.

2. Explain the various elements of decision making.

Ans : **(Imp.)**

Refer Unit-V, Page No. 147, Q.No. 4.

3. Define EOL. Explain the various steps for calculating EOL.

Ans : **(Imp.)**

Refer Unit-V, Page No. 149, Q.No. 5.

4. Briefly explain about the determination of EVPI and factors involved in the design of marketing research experiments.

Ans : **(Imp.)**

Refer Unit-V, Page No. 150, Q.No. 6.

5. Explain briefly the decision making under Uncertainty.

Ans : **(Imp.)**

Refer Unit-V, Page No. 159, Q.No. 10.

6. Define game theory. Explain the characteristics and objectives of game theory.

Ans : **(Imp.)**

Refer Unit-V, Page No. 162, Q.No. 13.

7. Define graphical method. Write down the algorithm for 2 × n, m × 2 games.

Ans : **(Imp.)**

Refer Unit-V, Page No. 167, Q.No. 23.

8. Define Linear Programming. Elucidate the Structure of LPP.

Ans : **(Imp.)**

Refer Unit-V, Page No. 179, Q.No. 24.

9. A company having a mechanical workshop has recently discontinued production of an unprofitable product. It has resulted in a considerable spare capacity. The company has decided to use this capacity to the maximum extent to produce three products which are profitable. The productivity coefficient in machine hours per unit and available machine time is given below:

Machine Type	Product 1	Product 2	Product 3	Time Available Machine Hours per week
Milling Machine	9	3	5	500
Lathe	5	4	0	350
Grinder	3	0	2	150

The sales department has indicated that the demand for Products 1 and 2 exceeds the maximum production rate whereas sales potential for Product 3 is 20 units per week. The profits for the three products have been estimated respectively as ` 3500, ` 1400 and ` 1750 for the three products. The company wants to decide the optimum level of production to maximize its profit.

Formulate this problem as a mathematical model.

Sol : **(Imp.)**

Refer Unit-V, Page No. 185, Prob. 13.

10. Describe the steps involved in graphical solution to linear programming models.

Ans : **(Imp.)**

Refer Unit-V, Page No. 186, Q.No. 31.

11. A company produces two types of pens A and B. Pen A is of superior quality and pen B of inferior quality. Profit on pen A and pen B are ` 5 and ` 3 per pen respectively. Raw material required for each pen A is twice as that of pen B. The supply of raw material is sufficient only for 1000 pens of B per day. Pen A requires a special clip and only 400 clips are available per day. For pen B only 700 clips are available per day. Find graphically the product mix so that the company can make maximum profit.

Sol : **(Imp.)**

Refer Unit-V, Page No. 186, Prob. 14.

12. The following table summarizes the key facts about two products, A and B, and resource Q, R and S, required to produce them.

Resource	Resource Usage Per Unit Produced		Amount of Resource Available
	Product A	Product B	
Q	2	1	2
R	1	2	2
S	3	3	4
Profit per unit	3	2	

Formulate the model and solve it graphically.

Sol : **(Imp.)**

Refer Unit-V, Page No. 188, Prob. 15.

UNIT I

Introduction

Research: Meaning, Purpose, Characteristics and Types - Process of Research: Formulation of objectives - Research plan and its components - Methods of Research: Survey, Observation, Case study, Experimental, Empirical and Comparative Methods - Difficulties in Business Research (Theory only)

Collection, Presentation & Analysis of data: Sources of Data: Primary and Secondary Sources - Methods of collecting Primary Data - Measurement and Scaling -Designing Questionnaires/Schedules. Census vs. Sampling - Methods of Sampling Random and Non-Random Sampling methods - Techniques. (Theory only)

1.1 RESEARCH

1.1.1 Meaning, Purpose

Q1. Define the term Research. Explain the purpose of Research.

(OR)

What is Research? State the purpose of research.

(OR)

What do you understand by the term Research?

(OR)

What do you mean by Research?

Ans : **(Imp.)**

Meaning

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

Research refers to search for knowledge. It can also be defined as a scientific and systematic search for gaining information and knowledge on specific topic or phenomena. Research is a systematic approach to gather information required for sound management decisions.

Research means;

i) Careful or thorough search.

ii) Serious inquiry or examination

iii) Collecting of information about a particular subject.

Definitions

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

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

Purpose of Research

i) Increasing knowledge within the discipline and

ii) Increasing knowledge within oneself as a professional consumer of research in order to evaluate and understand new developments within the discipline

iii) Increasing knowledge within the discipline can serve several sub purposes:

- To create methods to assess important concepts/phenomena
- To describe relevant phenomena and their relationships
- To provide evidence for the efficacy of a therapeutic technique or other change
- To provide support for the theoretical base of the discipline.

1.1.2 Characteristics

Q2. What are the characteristics of Research?

(OR)

State the various characteristics of Research.

Ans :

i) The research should focus on priority problems.

ii) The research should be systematic. It emphasizes that a researcher should employ a structured procedure.

iii) The research should be logical. Without manipulating ideas logically, the scientific researcher cannot make much progress in any investigation.

iv) The research should be reductive. This means that the findings of one researcher should be made available to other researchers to prevent them from repeating the same research.

v) The research should be replicable. This asserts that there should be scope to confirm the findings of previous research in a new environment and different settings with a new group of subjects or at a different point in time.

vi) The research should be generative. This is one of the valuable characteristics of research because answering one question leads to generating many other new questions.

vii) The research should be action-oriented. In other words, it should be aimed at reaching a solution leading to the implementation of its findings.

viii) The research should follow an integrated multidisciplinary approach, i.e., research approaches from more than one discipline are needed.

ix) The research should be participatory, involving all parties concerned (from policymakers down to community members) at all stages of the study.

x) The research must be relatively simple, timely, and time-bound, employing a comparatively simple design.

xi) The research must be as much cost-effective as possible.

xii) The results of the research should be presented in formats most useful for administrators, decision-makers, business managers, (or) the community members.

1.1.3 Types of Research

Q3. Explain different types of Research.

Ans : **(Imp.)**

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 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 uncon-trollable. One can only observe and report what is happening in a situation.

2. Analytical research

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.

3. Applied 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.

4. Fundamental research

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.

5. Quantitative 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.

6. Qualitative research

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.

7. Conceptual 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.

8. Empirical research

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.

1.2 PROCESS OF RESEARCH

Q4. Explain in detail the various stages of Research Process.

(OR)

Briefly describe the different steps involved in a Research Process.

(OR)

Explain briefly the process of Research.

Ans : **(Imp.)**

Research Process

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.

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.

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.

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 salepurchases 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 are those observations, which are collected by an investigator for the first time,

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

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.

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, acknowledgment and preface.

1.2.1 Formulation of Objectives

Q5. How to formulate objectives in a Research?

(OR)

Explain the purpose of Research Objectives.

Ans : **(Imp.)**

The objective of the research should be closely related to the research study of your dissertation. The main purpose of the research objective is to focus on research problem, avoid the collection of unnecessary data and provide direction to research study. Research is related to the aspiration and objectives are related to the battle-plan.

1. Specific

Objective should be clear and well defined. It helps to specify the research problems and provide proper guideline to find the solution of research problem. Specific objective identify the methods of collecting necessary information related to the research problem.

2. Measurable

Objectives should be measurable. It improves quality and quantitative of the research study to achieve its goal. The measurable research objectives provide guidelines for the improvement of research design. It is important element to achieve research objectives.

3. Achievable

Objectives should also be achievable in the time and it should provide accurate result from the use of sufficient resources in the specific time frame. It is related to effective measure of research problem. Achievable objectives ensure that every process of research is finished in accurate time will help to achieve the goals.

4. Realistic

Objective should be realistic, so that available resources like as men, money and machines could be used effectively. Objectives are most useful, when they accurately define the problem and take various steps that can be implemented with a specific time period.

5. Timely

Objective should be measured and achievable into the time frame. The research takes enough time in finding the solution of research problem. Timeline indicate when the objective will be accomplished.

1.3 RESEARCH PLAN

1.3.1 Components

Q6. What is Research Plan? Discuss the components of a Research Plan.

Ans : **(Imp.)**

Meaning

After identifying and defining the problem as also accomplishing the relating task, researcher must arrange his ideas in order and write them in the form of an experimental plan or what can be described as 'Research Plan'. This is essential specially for new researcher because of the following:

(a) It helps him to organize his ideas in a form whereby it will be possible for him to look for flaws and inadequacies, if any.

(b) It provides an inventory of what must be done and which materials have to be collected as a preliminary step.

(c) It is a document that can be given to others for comment.

Components

i) Research objective should be clearly stated in a line or two which tells exactly what it is that the researcher expects to do.

ii) The problem to be studied by researcher must be explicitly stated so that one may know what information is to be obtained for solving the problem.

iii) Each major concept which researcher wants to measure should be defined in operational terms in context of the research project.

iv) The plan should contain the method to be used in solving the problem. An overall description of the approach to be adopted is usually given and assumptions, if any, of the concerning method to be used are clearly mentioned in the research plan.

v) The plan must also state the details of the techniques to be adopted. For instance, if interview method is to be used, an account of the nature of the contemplated interview procedure should be given. Similarly, if tests are to be given, the conditions under which they are to be administered should be specified along with the nature of instruments to be used.

vi) A clear mention of the population to be studied should be made. If the study happens to be sample based, the research plan should state the sampling plan i.e., how the sample is to be identified.

vii) The plan must also contain the methods to be used in processing the data. Statistical and other methods to be used must be indicated in the plan. Such methods should not be left until the data have been collected. This part of the plan may be reviewed by experts in the field, for they can often suggest changes that result in substantial saving of time and effort.

1.4 METHODS OF RESEARCH

1.4.1 Survey, Observation, Case study, Experimental, Empirical and Comparative Methods

Q7. Explain the various methods of Research.

(OR)

Write about various methods of Research?

(OR)

Describe the various methods of Research.

Ans : **(Imp.)**

The following are the various methods of Research :

1. Survey

Survey is a 'fact finding' study. It is a method of research involving collection of data directly from a population or a sample at particular time. It must not be confused with the clerical routine of gathering and tabulating figures. It requires expert and imaginative planning, careful analysis and rational interpretation of the findings.

Data may be collected by observation, or interviewing or mailing questionnaires.

The analysis of data may be made by using simple or complex statistical techniques depending upon the objectives of the study.

2. Observation (or) Field Studies

Observation (or) field studies are scientific enquiries aimed at discovering the relations and interactions among sociological, psychological and educational variables in social institutions and actual life situations like communities, schools, factories, organizations and institutions. A social or institutional situation is selected and the relations among the attitudes, values, perceptions and behaviours of individuals and groups in the selected situation are studied.

3. Case study

A case study is an in-depth comprehensive study of a person, a social group, an episode, a process, a situation, a programme, a community , an institution or any other social unit.

It is one of the most popular types of research methods. Its purpose may be to understand the life cycle of the unit under study or the interaction between factors that explain the present status or the development over a period of time.

4. Experimental

Experimental research is designed to assess the effects of particular variables on a phenomenon by keeping the other variables constant or controlled. It aims at determining whether and in what manner variables are related to each other. The factor, which is influenced, by other factors is called a dependent variable, and the other factors, which influence it, are known as independent variables. For example, agricultural productivity, i.e., crop yield per hectare is a dependent variable and the factors such as soil fertility, irrigation, quality of seed, manuring and cultural practices which influence the yield are independent variables.

5. Empirical Research

Empirical research is defined as any research where conclusions of the study is strictly drawn from concretely empirical evidence, and therefore "verifiable" evidence.

This empirical evidence can be gathered using quantitative market research and qualitative market research methods.

6. Comparative Research

Comparative research, simply put, is the act of comparing two or more things with a view to discovering something about one or all of the things being compared. This technique often utilizes multiple disciplines in one study. When it comes to method, the majority agreement is that there is no methodology peculiar to comparative research. The multi disciplinary approach is good for the flexibility it offers.

1.5 Difficulties in Business Research

Q8. What are the difficulties involved in business research?

Ans : **(Imp.)**

In India, researchers in general, and business researchers in particular are facing several problems. This is all the more true in case of empirical research. Some of the important problems are as follows:

i) The lack of scientific training in the business research methodology is a major problem in our country. Many researchers take a leap in the dark without having a grip over research methodology. Systematic training in business research methodology is a necessity.

ii) There is paucity of competent researchers and research supervisors. As a result the research results many a time do not reflect the reality.

iii) Many of the business organizations are not research conscious and feel that investment in research is a wastage of resources and does not encourage research.

iv) The research and Development Department has become a common feature in many medium and large organizations. But decision makers do not appear to be very keen on implementing the findings of their R&D departments.

v) At the same time, small organizations which are the majority in our economy, are not able to afford a R & D department at all: Even engaging a consultant seems to be costly for them. Consequently, they do not take the help of research to solve their problems.

vi) Many people largely depend on customs, traditions and routine practices in their decision making, as they feel that research does not have any useful purpose to serve in the management of their Business..

vii) Even when research studies are undertaken, many a time, they are overlapping, resulting in duplication because there is no proper coordination between different departments of a university and between different universities.

viii) Difficulty of funds. Because of the scarcity of resources many university departments do not come forward to undertake research.

1.6 Sources of Data

1.6.1 Primary and Secondary Sources

Q9. What are the different types of Data Sources?

Ans : **(Imp.)**

Usually data is collected through two different sources i.e., primary and secondary and based on these sources data is classified into two types,

i) Primary data

ii) Secondary data.

i) Primary Data

Primary data refers to the data collected specifically which is for the purpose of research problem. It is the first hand information collected by the research firm or by an external agent with the objective of solving a research problem. There are different methods of collecting primary data. Researchers can conduct experiments to gather the required information. Other methods include questionnaires, mails, interviews of individuals, families, organizations, representatives etc.

ii) Secondary Data

Secondary data refers to the existing data that had been collected with an objective other than for research. It could be the data collected by the firm itself for any other purpose, or by any external party for the same or other research problem.

1.6.2 Methods of collecting Primary Data

Q10. Explain the methods of collecting primary data.

(OR)

Describe the various methods of collecting primary data.

(OR)

Critically examine the different methods of collecting primary data.

Ans : **(Dec.-18, Dec.-16, Dec.-13, Imp.)**

Meaning

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.

(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 :

i) Personal Interview

The investigator follows a rigid procedure and seeks answers to a set of preconceived 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.

ii) 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.

(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.

(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.

Q11. 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.

Q12. 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.

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.

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.

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.

1.7 MEASUREMENT AND SCALING

Q13. Define measurement and scaling. State the objectives to measurement and scaling.

Ans :

i) Measurement

According to W.S. Torgerson, "Measurement is the assignment of numbers to objects to represent amounts or degrees of a property possessed by all of the objects".

ii) Scaling

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, on a continuum. Such information helps the researcher in gaining the research objective of determining the stores image.

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

Objectives

There are two important objectives of measurement and scaling. First objective is to determine what can be revealed about the object of the study and second objective is to determine the appropriate statistic to be applied to analyses the information, following are the otte*-objectives of measurement and scaling,

1. Reliability

The measurement should be reliable by producing consistent results when certain research methods are repeated.

2. Validity

An instrument is valid when it measures what it is supposed to measure, but since attitudes are elusive, validity is not possible.

3. Sensitivity

The measurement should lead to identification of changes or differences.

4. Relevance

It Should facilitate the decision-maker by producing relevant information.

5. Versatility

Versatility or robustness of measurement for various statistical interpretations is a desired goal, especially for assessing the results validity and for drawing maximum meaning. Measurement should possess versatality.

Q14. Explain the properties of scales.

Ans :

Properties of Scales

The measurement scales has the following properties,

1. 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.

2. 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.

3. 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'.

4. 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.

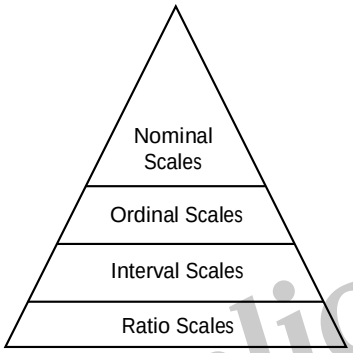
Q15. What are the different levels of measurement?

OR

Explain the various types of measurement scales with example.

Ans :

The following are the various levels/types of measurement scale,



1. Nominal Scaies

These scales involve labels. It is a qualitative scale, where the respondents answers are descriptive in nature like 'yes' or 'no', 'agree' or 'disagree' and so on. Examples of these scale are: designations as profession, region, religion, buyer/non buyer, smoker/ non-smoker, etc. It is a labelling scale that classifies or identifies the respondents, brands, attributes, etc. This scale measures raw responses and does not illustrate the intensity of responses.

It is to be noted that when a number is used for scaling, it does not reflect the quantity of the characteristic of an object/respondent. For instance, if the respondents are assigned code numbers as per their ID numbers, it does not mean that larger code number indicates the superiority of that respondent over the other.

2. Ordinal Scales

It is a ranking scale wherein, assigning numbers to objects imply the extent of presence of some characteristics. For e.g., students in a class are ranked 1 to 10 as per the marks they obtained. Here, 1st rank indicates the student's better performance over the student ranking 10th. It should be noted that the scale indicates only the position or relative size differences and not the magnitude of a respondent or object. In the above example, we know the student ranking 1st is better than the other students but we do not know how better is the student than the others.

3. Interval Scales

This scale measures the distance or the magnitude of difference in the ordinal scale. In our example of students, a scale of marks secured is an interval scale. Representing students on this scale shows as to which student is better and also helps in measuring their aptitude in terms of marks. For example, student ranking 1st is better than student ranking 2nd by 10 marks. Further, respondents or objects possessing a characteristic to the same extent would be placed in one scale. That is, students securing same marks would be on the same point of the scale.

4. Ratio Scales

Ratio scale possesses all the qualities of nominal, ordinal and interval scales inspite of its zero origin on the scale. With this scale, a researcher can classify, identify, rank, compare and compute the ratios of scale values for all the respondents/objects.

Q16. Distinguish between nominal, ordinal, Interval and Ratio Scale.

Ans : **(Imp.)**

Sl.No.	Basis	Nominal Scale	Ordinal Scale	Interval Scale	Ratio Scale
1.	**Meaning**	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 and Ratio scale.	Ratio scale helps the researchers to identify and rank the objects and compare the intervals or differences taking place between them.
2.	**Advantage/Use**	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.
3.	**Expression**	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.
4.	**Accuracy**	The accuracy of nominal scales can be measured with the help of chi-square and binominal test.	The accuracy of ordinal scales can he measured with the help of rank order correlation.	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 scales can be measured with the help of coefficient of variation.
5.	**Example of Usage**	Nominal scales are used for numbering the football players and social security.	Ordinal scales are used for ranking the teams in tournament, 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.

1.8 DESIGNING QUESTIONNAIRES/SCHEDULES

Q17. 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.

Types

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.

2. Non-Structured, Non-Disguised Questionnaire

The purpose of the study is clear, but the responses to the question are open-ended. 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".

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.

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 question naire, what comes out is "what does the respon dent know" rather than what they feels. Therefore, the endeavour in this method is to know the respondent's attitude.

Q18. What are the Components of Questionnaire.

Ans :

1. Words

The most obvious component is words. Researchers must carefully consider which words to use in creating the questions and scales for collecting raw data from respondents. The words selected by the researcher can influence respondent's answers to a given question.

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.

Q19. How to Organizing / Designing of Questionnaire?

Ans :

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.

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.

8. Decide on Layout and Reproduction

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

i) Acceptance of the Question- naire

ii) Ease of Control

iii) Ease of Handling

9. Check Questions

Once the first draft of the questionnaire has been completed, and before it is actually pretested, 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 pretested 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.

Q20. What are the Guidelines / Precautions for Preparation of Questionnaire ?

Ans : **(Imp.)**

The guidelines for preparing questionnaire are as follows :

1. Arrange Questions in a Logical Order

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

2. Design Items that Require Current and Easily Remembered Data

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

3. Questions should not be Ambiguous

The structure of the sentence as well as the word choice and order should not provide any room for misinterpretation on the part of the respondent.

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.

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.

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.

Q21. What are the Factors considered while Design of Questionnaire ?

Ans : **(Imp.)**

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.

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

i) Facts

ii) Quasifacts

iii) Awareness penetration of information

iv) Opinions

v) 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.

i) Open ended 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.

ii) Dichotomous questions

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 -

iii) Multiple choice questions

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.

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. C

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.

Q22. Define schedule. State the purpose of schedule.

Ans :

Meaning

Schedule is a popular tool used by the researchers for gathering the required information from the respondents. A schedule is similar to a questionnaire with only a difference that the questions contained in a schedule are filled up by the trained personnels who record/write down the responses. Such trained persons who record the respondent's answers are known as 'Enumerators.

Purpose

Some of the purposes of schedule are,

1. It is considered as a standardized tool for the purpose of interview or observation.
2. It assists in tabulation and analysis of data.
3. It is easily comparable because questions in schedule are standardized.
4. It act as a memory tickler.
5. Tabulation is easy due to structured form.

Q23. Distinguish between questionnaire and schedule.

Ans : **(Imp.)**

Some of the differences between questionnaire and schedule methods are,

Sl.No.	Area	Questionnaire	Schedule
1.	**Method of administration**	The questionnaire is sent via mail/ post to the target respondents /informants for filling it up.	The schedule method needs a research worker/ enumerator to fill up the answer from the respondent as they can interpret whenever needed.
2.	**Cost**	It is less expensive and economical as it involves only the cost of preparing questionnaires and mailing it to the respondents.	It is comparatively more expensive as it involves the cost of selecting enumerators, providing training to them, paying salaries and so on.
3.	**Rate of response**	The rate of response is comparatively.	A schedule method has a high response rate as the response is directly recorded by the enumerator.
4.	**Clarity and speed**	There is less accuracy present as answered the questionnaire who is unknown and not clear.	In case of a schedule, the identity of the person is known and therefore there is greater accuracy.
5.	**Rapport building**	Personal contact and rapport building is not and contact possible.	A direct personal contact can be established which helps in rapport building.
6.	**Scope of use**	Its scope is limited as it can be used only by the literate respondents.	It has a wider scope as information can be gathered from even illiterate respondents.
7.	**Population coverage**	Wider and greater representative distribution of sample is present in questionnaire method.	It is difficult to send the enumerators to distant places for collecting information.
8.	**Keep to success**	The quality and design of a question- naire acts as the key for a successful questionnaire.	The enumerators traits, such as intelligence, honesty, patience, sincerity and so on acts as the key to schedules success.
9.	**Use of observation**	Observations cannot be used under this method of collecting data.	While collecting data, the enumerators can make use of the observation method to judge the quality of respondents answers.

1.9 SAMPLING

Q24. Define Sampling. Explain the need of sampling.

(OR)

What do you understand by sampling?

Ans : **(Dec.-18)**

Meaning

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

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

Need

(i) Sampling can save time and money. A sample study is usually less expensive than a census study and produces results at a relatively faster speed.

(ii) Sampling is the only way when population contains infinitely many members.

(iii) Sampling may enable more accurate measurements for a sample study. It is generally conducted by trained and experienced investigators.

(iv) Sampling remains the only choice when a test involves the destruction of the items under study.

(v) Sampling usually enables to estimate the sampling error and thus, assists in obtaining information concerning some characteristics of the population.

Q25. What are the steps involved in sampling?

Ans :

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

1. Identifying the Target Population

This is the first step of sampling process. It involves collecting the element, which possesses the information collected by the researcher regarding which inferences are to be made. The target population studies can be well defined in terms of units of sampling, time and extent. The element can be an object from which the information can be obtained.

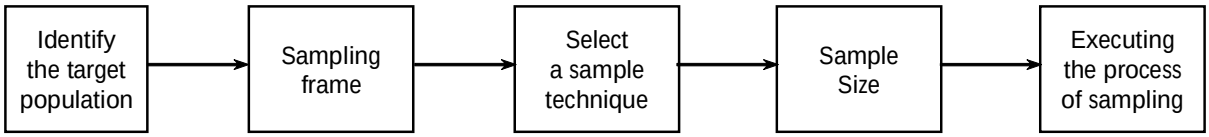


Fig.: Process of Sampling

2. Sampling Frame: The sampling frame represents the elements of target population. The sampling frame includes sets of direction for the identification of the target population.

3. **Selecting a Sampling Technique:** It includes decisions of broader nature. There are several approaches available to the researcher such as the Bayesian approach, Probability sampling or Non-Probability sampling approach.

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

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

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

1.9.1 Methods of Sampling

1.9.1.1 Random and Non-Random Sampling methods

Q26. Explain different types of Sampling Methods.

(OR)

Explain briefly the various random sampling methods.

Ans : **(Dec.-16, Imp)**

Sampling methods/techniques are broadly classified as,

(A) Random sampling methods

(B) Non-random sampling methods.

(A) Random Sampling Methods

This sampling process is random and the laws of probability are used in this sampling. Random sampling is of different types as follows,

1. Simple Random sampling
2. Stratified sampling
3. Systematic sampling
4. Multistage sampling.

1. Simple Random Sampling

In this technique, each and every item of the population is given an equal chance of being included in the sample. The selection is thus free from personal bias. This method is also known as the method of chance selection. It comprises of two methods. They are,

(i) Lottery methods

(ii) Table of random numbers.

(i) Lottery Methods: All items of the population are numbered or named on separate slips of paper of identical size and shape. The slips are then folded and mixed up in a container or drum. A random selection is then made of the number of slips required to constitute the desired size of sample.

(ii) Table of Random Numbers: Tippet's table of random numbers is popularly used. It consists of 10,400 four figure numbers. The digits in the table were chosen haphazardly and the Tippet's numbers have been subjected to numerous tests and are used in many investigations.

2. Stratified Sampling

This process divides the population into homogenous groups or classes called 'strata'. A sample is taken from each group by simple random method and the resulting sample is called a stratified sample.

A stratified sample may be either proportionate or disproportionate. In a proportionate stratified sampling plan, the number of items drawn from each stratum is proportional to the size of the strata. While in a disproportionate stratified sampling, equal number of items are taken from each stratum irrespective of the size of the stratum.

3. Systematic Sampling

This is used in those cases where a complete list of the population from which sampling is to be drawn is available. The method is used to select every K item from the list, where K refers to the sampling interval. The starting point is selected at random.

Example

If a complete list of 1000 students of a college is available and if a sample of 200 students is to be drawn, then every 5^{th} item (K = 5) must be taken. Suppose the straight point is 3 then the 1^{st} item is 3^{rd} student, second item would be 8^{th} students (3 + 5 = 8) the 3^{rd} item would be $l3^{th}$ student and so on.

4. Multistage Sampling

It refers to a sampling procedure, which is carried out in several stages. Suppose a sample of 10,000 house holds is to be drawn from the state of Telegana then in first stage, the states villages may be selected at random.

(B) Non-random Sampling Methods (Non-Probability Sampling)

A non-random sample is selected on the basis other than probability considerations as follows,

1. Judgement sampling
2. Quota sampling
3. Convenience sampling.

1. Judgement Sampling

In this method, the choice of sampling items exclusively depend upon the judgement of the investigations. In other words, the investigator exercise his judgements in the choice of sample items and includes those items in the sample which he thinks are most typical of the population with regards to the characteristics under investigation.

2. Quota Sampling

It is one of the type of judgement sampling. In this, quotas are set up according to the given criteria, but within the quotas the selection of sample items depends on personal judgement.

Example

Out of every 200 persons to be interviewed, 120 are housewives, 50 farmers, 30 children. Within the quota, the interviewer can select any people from the area concerned.

3. Convenience Sampling (Accidental Sampling)

Convenience sampling is also called 'chunk'. A chunk is a fraction of one population taken for investigation because of its convenient availability. A sample obtained from readily available lists such as telephone directories, automobile registrations is a convenient sample, even if the sample is drawn at random from the lists.

If the sample enter by 'accident' they just happen to be at the right place and at the right time. Therefore, it is also called accidental sampling.

Q27. What are the differences between Probability and Non-Probability Sampling

Ans :

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

Q28. What are the advantages of sampling ?

Ans :

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

1. Low cost of sampling

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

2. Less time consuming in sampling

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

3. Accuracy of data is high

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

4. Organization of convenience

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

5. Intensive and exhaustive data

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

6. Suitable in limited resources

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

7. Better rapport

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

Q29. What are the Applications of Sampling?

Ans : **(Imp.)**

Following are the Applications of Sampling.

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.

1.9.2 Census vs. Sampling

Q30. Distinguish between Census vs. Sampling.

Ans : **(Dec.-15)**

Sl.No.	Nature	Census	Sampling
1.	**Meaning**	A systematic method that collects and records the data about the members of the population is called Census.	Sampling refers to a portion of the population selected to represent the entire group, in all its Characteristics.
2.	**Enumeration**	Complete	Partial
3.	**Study of**	Each and every unit of the population.	Only a handful of units of the population.
4.	**Time required**	It is a time consuming	It is a fast process.

Short Questions and Answers

1. Research

Ans :

Meaning

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

Research refers to search for knowledge. It can also be defined as a scientific and systematic search for gaining information and knowledge on specific topic or phenomena. Research is a systematic approach to gather information required for sound management decisions.

2. Characteristics of Research.

Ans :

i) The research should focus on priority problems.

ii) The research should be systematic. It emphasizes that a researcher should employ a structured procedure.

iii) The research should be logical. Without manipulating ideas logically, the scientific researcher cannot make much progress in any investigation.

iv) The research should be reductive. This means that the findings of one researcher should be made available to other researchers to prevent them from repeating the same research.

v) The research should be replicable. This asserts that there should be scope to confirm the findings of previous research in a new environment and different settings with a new group of subjects or at a different point in time.

3. What is Research Plan?

Ans :

Meaning

After identifying and defining the problem as also accomplishing the relating task, researcher must arrange his ideas in order and write them in the form of an experimental plan or what can be described as 'Research Plan'. This is essential specially for new researcher because of the following:

(a) It helps him to organize his ideas in a form whereby it will be possible for him to look for flaws and inadequacies, if any.

(b) It provides an inventory of what must be done and which materials have to be collected as a preliminary step.

4. Primary Data

Ans :

Primary data refers to the data collected specifically which is for the purpose of research problem. It is the first hand information collected by the research firm or by an external agent with the objective of solving a research problem. There are different methods of collecting primary data. Researchers can conduct experiments to gather the required information. Other methods include questionnaires, mails, interviews of individuals, families, organizations, representatives etc.

5. Secondary Data

Ans :

Secondary data refers to the existing data that had been collected with an objective other than for research. It could be the data collected by the firm itself for any other purpose, or by any external party for the same or other research problem.

6. Disadvantages of primary data.

Ans :

i) 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.

ii) 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.

iii) 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.

7. Nominal Scaies

Ans :

These scales involve labels. It is a qualitative scale, where the respondents answers are descriptive in nature like 'yes' or 'no', 'agree' or 'disagree' and so on. Examples of these scale are: designations as profession, region, religion, buyer/non buyer, smoker/ non-smoker, etc. It is a labelling scale that classifies or identifies the respondents, brands, attributes, etc. This scale measures raw responses and does not illustrate the intensity of responses.

It is to be noted that when a number is used for scaling, it does not reflect the quantity of the characteristic of an object/respondent. For instance, if the respondents are assigned code numbers as per their ID numbers, it does not mean that larger code number indicates the superiority of that respondent over the other.

8. Ordinal Scales

Ans :

It is a ranking scale wherein, assigning numbers to objects imply the extent of presence of some characteristics. For e.g., students in a class are ranked 1 to 10 as per the marks they obtained. Here, 1st rank indicates the student's better performance over the student ranking 10th. It should be noted that the scale indicates only the position or relative size differences and not the magnitude of a respondent or object. In the above example, we know the student ranking 1st is better than the other students but we do not know how better is the student than the others.

9. Interval Scales

Ans :

This scale measures the distance or the magnitude of difference in the ordinal scale. In our example of students, a scale of marks secured is an interval scale. Representing students on this scale shows as to which student is better and also helps in measuring their aptitude in terms of marks. For example, student ranking 1st is better than student ranking 2nd by 10 marks. Further, respondents or objects possessing a characteristic to the same extent would be placed in one scale. That is, students securing same marks would be on the same point of the scale.

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.

Choose the Correct Answer

1. An image, perception or concept that is capable of measurement is called _____. [d]
   (a) Scale (b) Hypothesis
   (c) Type (d) Variable
2. In order to pursue the research, which of the following is priorly required? [b]
   (a) Developing a research design (b) Formulating a research question
   (c) Deciding about the data analysis procedure (d) Formulating a research hypothesis
3. What are the conditions in which Type-I error occurs? [b]
   (a) The null hypotheses get accepted even if it is false
   (b) The null hypotheses get rejected even if it is true
   (c) Both the null hypotheses as well as alternative hypotheses are rejected
   (d) None of the above
4. 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
5. Which of the following is not the method of Research? [c]
   (a) Survey (b) Historical
   (c) Observation (d) Philosophical
6. 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
7. 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
8. The main objective of _____ study's to acquire knowledge [b]
   (a) Exploratory (b) Descriptive
   (c) Diagnostic (d) Descriptive and Diagnostic
9. _____ 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
10. One of the terms given below is defined as a bundle of meanings or characteristics associated with certain events, objects, conditions, situations, and the like. [c]
   (a) Construct (b) Definition
   (c) Concept (d) Variable

Fill in the blanks

1. _____ is a systematized effort to gain new knowledge.
2. _____ research basically describes what is. It mainly involves collection, recording, describing and analyzing the facts related to the study.
3. _____ research is action oriented or solution oriented.
4. _____ research is based on quantitative variables, which can be measured in appropriate units.
5. _____ research involves the development of new theories, abstract ideas, and generalized principles.
6. _____ is a search for knowledge.
7. _____ is the first and foremost step in a research process.
8. A _____ is a well-defined plan of action.
9. A _____ is an in-depth comprehensive study of a person, a social group, an episode, a process, a situation, a programme, a community, an institution or any other social unit.
10. According to W.S. Torgerson, _____ is the assignment of numbers to objects to represent amounts or degrees of a property possessed by all of the objects.

ANSWERS

1. Research
2. Descriptive
3. Applied
4. Quantitative
5. Conceptual
6. Research
7. Formulation of a problem
8. Research design
9. Case study
10. Measurement

One Mark Answers

1. Survey

Ans :

Survey is a 'fact finding' study. It is a method of research involving collection of data directly from a population or a sample at particular time.

2. Scaling

Ans :

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, on a continuum.

3. Ratio Scales

Ans :

Ratio scale possesses all the qualities of nominal, ordinal and interval scales inspite of its zero origin on the scale.

4. Structured, Non-Disguised Questionnaire

Ans :

Most questionnaire studies made in marketing research are of the first type- they are structured and are not disguised. If the sales manager for a musical instrument company wants to find out how many and what type of people play various types of instruments, a formal list of questions may be set up that asks directly about the ownership and playing of various instruments.

5. Stratified Sampling

Ans :

This process divides the population into homogenous groups or classes called 'strata'. A sample is taken from each group by simple random method and the resulting sample is called a stratified sample.

UNIT II

INTERPRETATION AND REPORT WRITING:

Interpretation: Introduction - Essentials for Interpretation, Precautions in interpretation - Conclusions and generalization - Methods of generalization. Statistical fallacies: bias, inconsistency in definitions, inappropriate comparisons, faulty generalizations, drawing wrong inferences, misuse of statistical tools, failure to comprehend the data. (including small cases).

Report Writing: Meaning and types of reports - Stages in preparation of Report - Characteristics of a good report - Structure of the report'-Documentation: Footnotes and Bibliography - Checklist for the report.

2.1 INTERPRETATION

2.1.1 Introduction, Essentials for Interpretation

Q1. Define Interpretation.

(OR)

What is Interpretation of Data? What are the Essentials for Interpretation?

(OR)

Why is Interpretation considered a basic component of research process.

Ans : **(Dec.-18, Dec.-16, Dec.-14, Imp.)**

Meaning

Interpretation refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study. In fact, it is a search for broader meaning of research findings. The task of interpretation has two major aspects viz.,

(i) The effort to establish continuity in research through linking the results of a given study with those of another, and

(ii) The establishment of some explanatory concepts. Interpretation is concerned with relationships within the collected data and the extension of study beyond the collected data as well.

Definitions

- "The task of drawing conclusions or inferences and of explaining their significance after a careful analysis of selected data is known as interpretation".
- "It is an inductive process, in which you make generalizations based on the connections and common aspects among the categories and patterns".

Certain points should be kept in mind before proceeding to draw conclusions from statistics. It is essential that:

(a) The data are homogeneous

It is necessary to ascertain that the data are strictly comparable. We must be careful to compare the like with the like and not with the unlike.

(b) The data are adequate

Sometimes it happens that the data are incomplete or insufficient and it is neither possible to analyze them scientifically nor is it possible to draw any inference from them. Such data must be completed first.

(c) The data are suitable

Before considering the data for interpretation, the researcher must confirm the required degree of suitability of the data. Inappropriate data are like no data. Hence, no conclusion is possible with unsuitable data.

(d) The data are properly classified and tabulated

Every care is to be taken as a pre-requisite, to base all types of interpretations on systematically classified and properly tabulated data and information.

(e) The data are scientifically analyzed

Before drawing conclusions, it is necessary to analyze the data by applying scientific methods. Wrong analysis can play havoc with even the most carefully collected data.

2.1.2 Precautions in Interpretation

Q2. Explain the precautions to be undertaken in interpretation of data.

(OR)

What precautions are necessary in interpretation of data.

Ans : **(Dec.-14. Dec.-13)**

Researcher must pay attention to the following points for correct interpretation:

(i) The data are appropriate, trustworthy and adequate for drawing inferences;

(ii) The data reflect good homogeneity;

(iii) Proper analysis has been done through statistical methods.

(iv) The researcher must remain cautious about the errors that can possibly arise in the process of interpreting results. He should be well equipped with and must know the correct use of statistical measures for drawing inferences concerning his study.

(v) Interpretation is intertwined with analysis and cannot be distinctly separated.

(vi) He must never lose sight of the fact that his task is not only to make sensitive observations of relevant occurrences, but also to identify and disengage the factors that are initially hidden to the eye.

2.2 Conclusions and Generalization

Q3. Explain briefly about Conclusions and generalization.

Ans :

A conclusion is a finding drawn from a set of data in a study (or) experiment.

Results are direct observations summarized and integrated by the statistical analysis such as comparison of two groups of workers. Group 'A's' average wage is Rs. 5,000 and that of group 'B' is Rs. 6,000. A conclusion is an inference based on the data that group B workers are better paid than those of group 'A'.

In every day life, we often make generalizations. We believe that what is true of the observed instances will be true of the unobserved instances. Since, we have had an uniform experience, we expect that we shall have it even in the future. We are quite conscious of the fact that the observed instances do not constitute all the members of a class concerned. But we have a tendency to generalize. A generalization is a statement, the scope of which is wider than the available evidence.

For example: A is a crow, it is black. B is a crow, it is black. C is a crow, it is also black. Therefore, it can be generalized that "all crows are black". Similarly, all swans are white. All rose plants possess thorns etc., The process by which such generalizations are made is known as induction by simple enumeration.

2.3 Methods of Generalization

Q4. Discuss the methods of generalization.

(OR)

Explain the various methods of generalization.

Ans : **(Dec.-15, Imp.)**

Normally, two methods are used for generalization viz.

1. Logical method and
2. Statistical method

There are other methods also for generalization, but generally only these two methods are widely used. Let us discuss these two methods in detail.

1. Logical Method

This method was first introduced by John Stuart Mill, who said that generalization should be based on logical processes. Mill thought that discovering causal connections is the fundamental task in generalization. If casual connections hold good, generalization can be done with confidence.

Five methods of experimental enquiry have been given by Mill. These methods serve the purpose of discovering causal connections.

These methods are as follows.

(i) The Method of Agreement

This may be positive or negative. The method of agreement states that if two or more instances of a phenomenon under investigation have only one circumstance in common, the circumstance is the cause or the effect of the given phenomenon. For

example, a person gets pain in his eyes whenever he roams in the sun. Negatively, when he is under the shade he does not have pain. Therefore, the cause for pain is roaming in the sun.

(ii) The Method of Difference

This method is a combination of both positive and negative methods of agreement. In this method only two instances are required. The two instances resemble each other in every other respect, but differ in the absence or presence of the phenomenon observed. The circumstance in which alone the two instances differ, is the effect, or the cause.

Let us take the example given by Mill. A man is shot, he is wounded and dies. Here the wound is the only differentiating circumstance between the man who is alive and the man who is dead. Hence, death is caused by the wound.

(iii) Joint Method of Agreement and Difference

This is a combination of the method of agreement and the method of difference. According to this method, we require two sets of instances.

If two or more instances in which the phenomenon occurs have only one circumstance in common, while two or more instances in which it does not occur have nothing in common, save the absence of that circumstance, the circumstance in which alone the two sets of instances differ, is the effect or the cause.

For example :

A + B + C Produce X

A + P + Q Produce X

M + N + Non-A Produce Non-X

G + H + Non-A Produce Non-X

∴ A and X are causally connected.

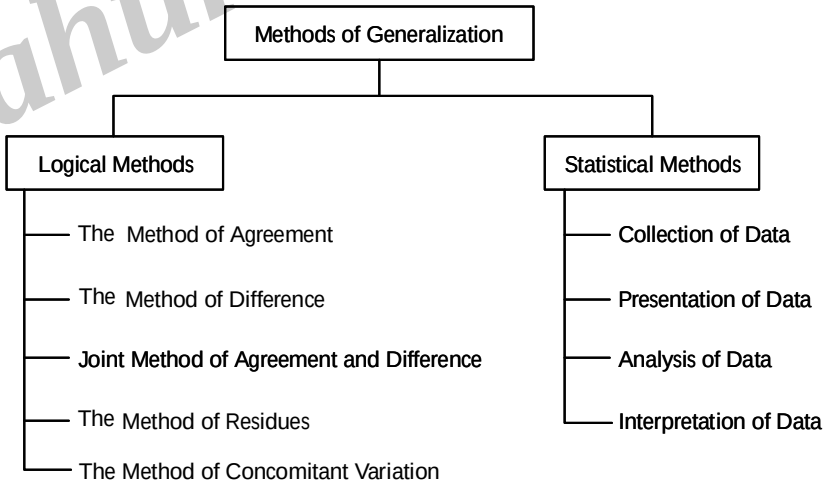


Fig.: Methods of generalization

(iv) The Method of Residues

This method is based on the principle of elimination. The statement of this method is that, subtract from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents.

For example: A loaded lorry weighs 11 tons. The dead weight of the lorry is 1 ton. The weight of load = 11 – 1 = 10 tons.

(v) The Method of Concomitant Variation

This method can be stated as "whatever phenomenon varies in any manner, whenever another phenomenon varies in some particular manner, is either the cause or the effect of that phenomenon or is connected with it. This method is quantitative in nature and needs statistical techniques for measurement. That is why it is also known as the method of quantitative induction, because we base our inference on quantitative change in the two factors and is applied as some form of correlation analysis.

2. Statistical Method

Statistical method may be defined as " the collection, presentation, analysis and interpretation of numerical data". Thus statistical method involves four steps:

(i) Collection of Data

The facts pertaining to the problem under study are to be collected either by survey method or by observation method or by experiment or from a library.

(ii) Presentation of Data

The data collected as to be processed by classification, tabulation and then be presented in a clear manner.

(iii) Analysis of Data

The processed data then should be properly analyzed with the help of statistical tools, such as measures of central tendency, measures of variation, measures of sknewness, correlation, time series, index numbers etc.

(iv) Interpretation of Data

The collected and analyzed data has to be interpreted. It involves explanation of facts and figures and drawing inferences and conclusions.

2.4 Statistical Fallacies

2.4.1 Bias, Inconsistency in definitions, Inappropriate comparisons, Faulty generalizations, Drawing wrong inferences, Misuse of statistical tools, Failure to comprehend the data

Q5. Define Statistical Fallacies.

(OR)

State the different statistical fallacies that may arise while interpreting the data. Explain each one of them with relevant examples.

(OR)

Discuss how the statistical fallacies arise while interpreting the data. Explain with suitable examples.

(OR)

What are the statistical fallacies and how they effects the results.

Ans : **(Dec.-18, Dec.-16, June-16, Dec.-14, Dec.-13, Imp.)**

Interpretation of data, is a very difficult task and requires a high degree of care, objectivity, skill and judgement. In the absence of these things, it is likely that the data may be misused. In fact, experience shows that the largest number of mistakes are committed knowingly or unknowingly while interpreting statistical data which may lead to misinterpretation of data by most of the readers.

Statistical fallacies may arise at any stage – in the collection, presentation, analysis and interpretation of data. The following are some of the

(i) Specific examples illustrating how statistics can be misinterpreted,

(ii) Sources of errors leading to false generalizations,

(iii) Examples how fallacies arise in using statistical data and statistical methods.

1. Bias

Bias is very common in statistical work and often leads to false conclusions. For example, if a researcher wants to show that the wages at a particular company are very low he may select the sample in such as manner as to exclude high paid workers. Similarly, businessmen may use

statistics to prove the superiority of their own business over others.

For example, a business owner may present data from a quarter in which his business performed relatively better compared to the competition. Of course, the conclusion would only be valid if the data included a representative number of quarterly revenue.

2. Inconsistency in Definitions

Some times false conclusions are drawn because of failure to define properly the object being studied and hold that definition in mid for making comparisons. When the working capital of two firms is compared, net working capital of one must be compared with only net working capital of the other and not with gross working capital. Even within the organization, for facilitating comparison over a period of time it is necessary to keep the definition constant.

3. Inappropriate Comparisons

Comparisons between two things can not be made unless they are really alike. Unfortunately, this point is generally forgotten and comparisons are made between two dissimilar things, thereby, leading to fallacious conclusions.

4. Faulty Generalizations

Many a time people jump to conclusions (or) generalizations on the basis of either too small a sample or a sample that is not representative of the population.

5. Drawing Wrong Inferences

Some times wrong inferences may be drawn from the data. For example, the population of a town has doubled in 10 years. From this it is interpreted that the birth rate in the town has doubled. Obviously, this is a wrong inference, as the population of the town can double in many ways (example: exodus from villages, migration from other places etc.) than doubling of birth rate only.

6. Misuse of Statistical Tools

The various tools of analysis such as measures of central tendency, measures of variation, measures of correlation, ratios, percentages etc., are very often misused to present information in such a manner as to convince the public or to camouflage things. In a company there are 1,00,000 shares and 1,000 share holders. The company claims that their shares are well distributed as the average share holding is 100. But a close scrutiny reveals that 10 persons hold 90,000 shares where as 990 persons hold 10,000 shares, average being about 10. Similarly, range can be misused to exaggerate disparities.

7. Failure to Comprehend the Data

Very often figures are interpreted without comprehending the total background of the data and it may lead to wrong conclusions. For example, see the following interpretations:

- The death rate in the army is 9 per thousand, where as in the city of Delhi it is 15 per thousand. Therefore, it is safer to be in the army than in the city.
- Most of the patients who were admitted in the intensive care (IC) ward of a hospital died. Therefore, it is unsafe to be admitted to intensive care ward in that hospital.

2.5 REPORT WRITING

2.5.1 Meaning

Q6. What is a Report and Report writing. State its objectives.

Ans : **(Dec.-14, Imp.)**

(i) Reports

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, planning future course of action and taking effective decisions. Reports acts as a tool for providing feedback to the managers related to various aspects of the organization.

Definition

The compilation of information which has been sought out, collected sifted, organized and written to convey a message is usually referred as a 'report'.

(ii) Report Writing

Report writing is a conscious, rational and systematic effort. It is both an art as well as science. Report-writing requires conceptual and communication skills and a scientific approach to investigation, analysis and presentation. Managers should possess right abilities and attitudes for creating effective reports.

Objectives

Reports are written for various purposes such as,

(i) For reviewing performance

(ii) Keep a check on a continuing activity

(iii) Survey the market

(iv) Future planning needs of the organization

(v) Submit standardized information etc.

2.5.2 Types of Reports

Q7. Explain in detail the various types of Reports.

(OR)

Classify the different types of reports.

Ans :

The following figure depicts the various types of reports.

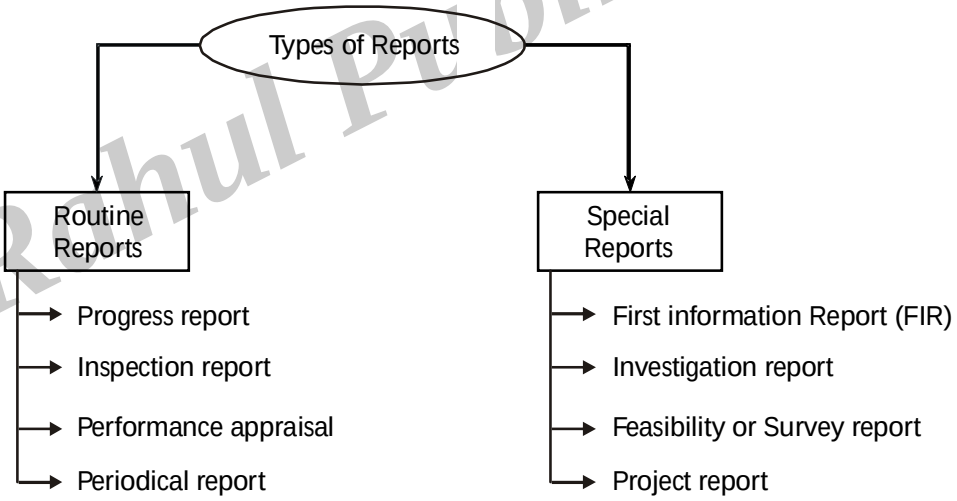


Fig. : Types of Reports

1. Routine Reports

Routine reports consists of forms in which blanks need to be filled in or multiple choice statements are to be ticked. Report forms are prepared efficiently because it gives complete information to management. The different types of routine reports are,

(a) Progress Report

Progress report provides information regarding progress of a project during completion of the project like manufacture of products or implementation of a scheme, construction of a building etc.

(b) Inspection Report

After the completion of inspection, inspection report is submitted. Inspection is necessary for determining irregularities or deviation from standard practice in the daily activities of the company.

Example: Internal audit reports.

Inspection of machinery, buildings and property are conducted at periodic intervals. Audits and inspection of work are done to know whether the work is carried out properly or not.

(c) Performance Appraisal

Performance appraisal is prepared by filling a form periodically. It helps in assessing the performance of an employee. Supervisor job is to fill an assessment report regarding the working of the subordinates. Performance appraisal thus help superiors to assess the performance and to provide feedback on their performance. Performance appraisal reports helps in taking decisions related to promotions and other employee benefits.

(d) Periodical Report

Periodical reports are made at regular intervals and provides information about the day-to-day work. These reports are usually prepared by filling in a form.

2. Special Reports

(a) First Information Report (FIR)

This report is always written by an individual. This report is needed for reasons including disaster like fire, building collapse, robbery or accident in an organization. It is prepared soon after the incident occurs and is submitted to a higher authority to take necessary action. The report should give information like what had happened, about what time, who first noticed and what steps were taken to that incident. It also provide information related to the extent of destruction or loss of property and important documents by fire etc.

(b) Investigation Report

This report is usually prepared after making thorough enquiry. Investigation is done when a problem rises and management find alternatives to solve those problems. Investigation is usually done for the cases like falling sales, declining deposits in a bank, customer complaints, losses in a branch etc.

Investigation requires lot of thinking, discussion and consideration for arriving to a solution. Investigation related to serious problems are done by competent and experienced persons.

(c) Feasibility or Survey Report

This type of report is prepared when an organization launches a new product in the market, introduce a new service or any major changes which have its impact on company's customer.

The factors to be examined totally depends on the purpose of the survey. The purpose of conducting this report may be due to the factors like suitability of a site for a factory, to evaluate the feasibility and financial viability of a proposal, to survey the market, to estimate damage etc.

(d) Project Report

This report is written after the completion of the preliminary survey and shaping up of a new proposal. It indicates the cash flow and expected results for the proposal undertaken. This report facilitates planning and convincing sanctioning and funding authorities like government, departments and banks.

2.5.3 Stages in preparation of Report

Q8. Discuss briefly the stages in Report Preparation.

(OR)

Explain the various stages in preparation of report.

Ans : **(June-16, Dec.-14, Imp.)**

Research reports are the product of slow and painstaking and accurate work.

Therefore, the preparation of the report may be viewed in the following major stages.

(1) The logical understanding and analysis of the subject matter.

(2) Planning/designing the final outline of the report.

(3) Write up/preparation of rough draft.

(4) Polishing/finalization of the Report.

1. Logical Understanding of the Subject Matter

It is the first stage which is primarily concerned with the development of a subject. There are two ways to develop a subject viz. a. logically and b. chronologically. The logical development is done on the basis of mental connections and associations between one aspect and another by means of logical analysis. Logical treatment often consists of developing material from the simple to the most complex.

Chronological development is based on a connection or sequence in time or happening of the events. The directions for doing something usually follow the chronological order.

2. Designing the Final Outline of the Report

It is the second stage in writing the report. Having understood the subject matter, the next stage is structuring the report and ordering the parts and sketching them. This stage can also be called as planning and organization stage. Ideas may pass through the author's mind.

Unless he first makes his plan/sketch/design he will be unable to achieve a harmonious succession and will not even know where to begin and how to end. Better communication of research results is partly a matter of language but mostly a matter of planning and organizing the report.

3. Preparation of the Rough Draft

The third stage is the write up/drafting of the report. This is the most crucial stage to the researcher, as he/she now sits to write down what he/she has done in his/her research study and what and how he/she wants to communicate the same.

4. Finalization of the Report

This is the last stage, perhaps the most difficult stage of all formal writing. It is easy to build the structure, but it takes more time for polishing and giving finishing touches. Take for example the construction of a house. Up to roofing (structure) stage the work is very quick but by the time the building is ready, it takes up a lot of time.

2.5.4 Characteristics of a Good Report

Q9. Describe the Characteristics of a Good Report.

Ans : **(June-16)**

Several characteristics of a good report includes

1. Precision

In a good report, the report writer is very clear about the exact and definite purpose of writing the report. This central purpose directs the investigation, analysis, recommendations, and others. The precision of a report provides unity to the report and makes it a valuable document for best usage.

2. Accuracy of Facts

Information contained in a report must be based on accurate facts. Since decisions are based on report information, inaccurate information or statistics will lead to a wrong decision. It will hamper ensuring the achievement of the organizational goal.

3. Relevancy

The facts presented in a report should be accurate and relevant. Irrelevant facts make a report confusing and likely to be misleading.

4. Simple Language

Simplicity is the best for anything. It is just another essential feature of a good report. A good report is written in simple language, avoiding vague and unclear words. The writer's emotion or goal should not influence the report's language. The message of a good report should be self-explanatory. A good reporter should use simple sentences instead of complex sentences in the narration of facts.

5. Conciseness

A good report should be concise, but it does not mean that a report can never belong. Rather it means that a report transmits maximum information with minimum words. It avoids unnecessary detail and includes everything significant and necessary to present proper information.

6. Grammatical

A good report is free from errors. Any faulty construction of a sentence may change its meaning in the reader's mind and potentially make it confusing or ambiguous.

7. Unbiased Recommendation

Recommendations usually affect the reader's mind. Therefore, if recommendations are made at the end of a report, they should be impartial and objective and come to a logical conclusion for investigation and analysis.

8. Clarity

Clarity depends on the proper arrangement of facts. A good report is absolutely clear. The reporter should clarify their purpose, define their sources, state their findings, and make a necessary recommendation A report must be clear to understand to ensure effective communication throughout.

9. Presentation

The presentation of a report is also a factor that should be considered for a good report. The structure, content, language, typing, and presentation style of a good report should be attractive to make a clear impression in the mind of its reader.

10. Complete Information

A good report shows important information. Most of this information is analyzed as the basis of importance. The report should not contain useless or vague information.

2.5.5 Structure of the Report

Q10. Discuss the structure of a Report.

(OR)

Out line the structures of a Research Report.

Ans : **(Dec.-18, Dec.-13, Imp.)**

Report format includes the following items they are,

1. Cover page
2. Introductory page
   - (a) Vita
   - (b) Acknowledgment
   - (c) Abstract
   - (d) Table of contents
   - (e) List of tables
   - (f) List of figures
   - (g) List of variables.
3. Text
   - (a) Chapters
   - (b) Footnotes
   - (c) Conclusions.
4. Bibliography
5. Appendices.

1. Cover Page

The cover page includes the title of the research, department to which reports are submitted, researcher's name, supervisor's name, university emblem, university name and its address, month and year of submission etc.

2. Introductory Pages

The items of the introductory pages can be explained as follows.

(a) Vita

It gives a brief information about the researcher. It comprises name, date of birth, place of birth, educational qualification, work experience and list of publications by the researcher.

(b) Acknowledgement

During research process, the people who helped, suggested, criticized should be acknowledged in the report.

(c) Abstract

It gives brief explanation of the research report outlining the title of the research, objectives, limitations of present literatures, the requirement for further research, proposed performance measure and research methods, outcomes of comparisons highlights of outcomes and a brief explanation of case study.

(d) Table of Contents

If the report is lengthy or if it is divided into numerous parts, it is usually desirable to have a table of contents.

(e) List of Tables

For data and results presentation, tables are used in research reports Every table needs to be numbered in such a manner that it is internal to each chapter.

(f) List of Figures

The list of figures are used to represent results in research report in order to have better understanding and comparison. Like tables, all figures should also be numbered in such a manner that it is internal to each chapter.

(g) List of Variables

Different variables and abbreviations are used in different chapters of report to explain about different models, equations and results.

3. Text

The text includes various items in a research report such as-chapters, main sections, sub-sections, footnotes, conclusions etc.

(a) Chapters

Based on the type of research, the text of the report needs to divided into different chapters. The different chapters are as follows,

(i) Introduction

This section serves to introduce the reader to the research project. It should give the background of the problem, the importance of the problem, the various dimensions of the problem and whether any previous research was done which is pertinent to the specific project being reported.

(ii) Literature Review

A detailed literature review of the research topic must be carried out. With the help of reading and reviewing the previous research data, about the topic, the researcher gets a better understanding about the topic. It would be helpful for him in setting up a meaningful research objective.

(iii) Objective and hypothesis

The researcher should clearly describe the objectives and related hypothesis based on literature review.

(iv) Research Methodology

Research methodology refers to the system of models, procedures and techniques which are used to determine the results of a research problem.

(v) Data Analysis

Prof. Wilknison and Bhandarkar states, analysis of data basically comprises various closely related activities which are carried out with an aim to outline the collected data and organizing it in such a way that it would help in answering the research questions or suggest hypothesis or questions. If in case no such questions or hypothesis has started the study.

(b) Footnotes

Footnotes are indicated with various symbols, notes at the end of the page in a printed document.

(c) Conclusions

In conclusion, the chapter briefly discuses about research findings, contribution of the study, limitations and scope of the study.

4. Bibliography

It refers to the list of references which are related to a topic or subject. It includes all the information which is found in the first footnote and it is related to the research work. It is present at the end of the main body of a report. Bibliography differs in its functions from that of the footnotes.

The main function of the bibliography is to list in an alphabetical order all the references which are being used by the researcher.

5. Appendices

The purpose of the appendix is to provide a place for those report items which do not fit in the research report properly because they are either too detailed or are too specific.

2.5.6 Documentation

2.5.6.1 Footnotes

Q11. Write a brief note on Footnotes in the context of a Research Report.

(OR)

Define and discuss footnotes in Research Report. Explain the purposes of footnotes.

Ans : **(Dec.-18, Dec.-16, Imp.)**

Meaning

Footnotes are notations (notes) at the end of the page in a printed document. They are presented in a smaller font size traits the dominant text. They are indicated with superscript numbers or various symbols. When many footnotes are present in a text, numbers are used to indicate them and when few footnotes are present, symbols are used to indicate them. The asterisk (*) symbol is most commonly used for footnotes.

The use of footnotes in the research report helps in servicing the two important objectives first the recognition of the source of materials used in quotations are made so that the reader can easily cross verily any references. Further footnotes provide supplemental values to the report.

The researcher needs to take into consideration the following points with reference to the footnotes,

(i) A footnotes may be separated from the textual material by a line, which is spaced half an inch and a line about one and a half inches long from the textual material.

(ii) All the footnotes must be numbered in a sequential order usually starting with 1 in each chapter separately.

(iii) In case of statistical tables or other numerical material, an asterisk symbol (*) or a similar type of a symbol may be used in order to avoid confusion.

(iv) Footnotes must always, be typed in single space eventhough they are divided from one another by double space.

Purposes

Footnotes are used for a variety of purposes as follows,

1. They are commonly used to provide citations to reliable sources. Footnoting citations helps to keep the flow of prose intact and this in turn holds the concentration of the reader on the text.
2. They are used to provide explanatory or additional information pertaining to a topic in the text. This additional information can be of interest to the reader.
3. They are used to make amusing comments and side notes which would be inappropriate in the central text.
4. They are sometimes also used to provide subtle commentary on a topic or source and to recount anecdotes (short interesting real story).

2.5.6.2 Bibliography

Q12. What is Bibliography? What are the various sources of Bibliography?

Ans :

Meaning

It refers to the list of references which are related to a topic or subject. It includes all the information which is found in the first footnote and it is related to the research work. It is present at the end of the main body of a report. Bibliography differs in its functions from that of the footnotes.

The main function of the bibliography is to list in an alphabetical order all the references which are being used by the researcher. But the foot notes only provides the identification details of a work.

The format of a research bibliography is similar to that of the format of a footnote except a single difference i.e., in bibliography, no page number is mentioned in the case of the book/reports enlisted in the bibliography.

Sources

Usually the important sources of bibliography includes the following,

(a) Research books, and articles published in research journals or periodicals.

(b) Annual research reports which are issued by government or non-government organizations.

(c) Research information collected by inter viewing the research population.

(d) Research related information collected from multimedia sources like, Radio and Television talks.

(e) The information collected from attending the lectures.

2.5.6.3 Checklist for the Report

Q13. Explain the Checklist for the Research Report.

Ans : **(Imp.)**

1. Letter of transmittal

- Follow formatting conventions
- Include all necessary elements, including the return address, date, inside address, salutation, subject line
- Introduce the report by name?
- Acknowledge any assistance received from the recipient of the letter?
- Briefly highlight key findings, conclusions, and recommendations from the report?
- Maintain a professional tone?

2. Title page

- Include an informative title that indicates the focus and nature of the report (e.g., evaluation, recommendation, feasibility study, survey, laboratory report)?
- Include the date and the author's name, title, and organizational affiliation?
- Indicate for whom the report was written?

3. Table of contents

- List all first- and second-level section headings (and, optionally, third-level headings)?
- Reproduce the headings and numbering exactly from the body of the report?
- Include the full titles of all appendices?

4. List of figures and the list of tables

- List all figures and tables (under separate headings), giving page numbers for each?
- Reproduce exactly the numbering and full titles of each figure and table?

5. The glossary

- Present specialized terms in alphabetical order?
- Define terms in a grammatically parallel way (e.g., all in sentences or all in phrases)?

  Note: That integrating definitions into the body of the report is generally a good strategy.

6. Does the executive summary

- Summarize the report in such a way that it could stand on its own and would make sense to a managerial, public, or non-technical audience?
- Stay within one or two pages (preferably one) and within a 500-word limit?
- Briefly explain the focus, context, and research method for the report?
- Summarize the key information from the report, concentrating on the problem, conclusions, recommendations, and financial implications (if relevant)?

7. Introduction section

- Explicitly introduce the report and briefly explain the focus and context for the report

  Note: Note that the introduction may be followed by related sections or subsections, including

- **Objectives of the study** (often formatted as a list of "To…" points)
- **Scope** (indicating what your analysis includes and excludes)
- **Background** (e.g., about the problem or organization being discussed)
- **Review of research (literature review)**

8. Research method section

- Explain when, where, and how the research was done?

- Describe your research method in sufficient detail?
- Include informative subheadings if appropriate

9. Body of your report

- Present, interpret, and discuss your findings in a series of sections with informative headings and subheadings
- Present findings clearly and coherently, with a brief introduction to each section?
- Use lists, tables, and figures (e.g., graphs) to present information where appropriate?
- Refrain from presenting conclusions and recommendations?
- Include properly formatted citations for all information taken from sources?

10. Figures and Tables

- Numbered in separate series (e.g., Table 1, Table 2, Figure 1, Table 3, Figure 2)?
- Informatively titled (e.g., Table 1**:** Annual precipitation in Calgary, 1980 - 1990)?
- Properly formatted, with the title for tables above and that for figures below?
- Accompanied by legends or footnotes to explain abbreviations or provide information necessary to interpret the information presented?
- Referred to and discussed in the body of the report?
- Accompanied by a source citation, if the information was borrowed or adapted?

11. Conclusions Section

- Summarize the most important findings in the report and indicate their significance
- Include only conclusions that flow from the data and discussion presented?
- Show how the report's objectives have been met?
- Refrain from introducing new information?

12. Does the recommendations section

- Suggest actions that should be taken or considered in light of the report's conclusions?
- Present recommendations in a numbered list if there are several

13. Writing style generally characterized by

- Sensitivity to the audience
- Paragraphs that begin with a topic sentence
- Straight forward, concise, and natural wording
- Precise wording
- The use of strong verbs and active voice
- Grammatical parallelism in lists and headings
- Relatively few sentences over 30 words or two typed lines?
- Relatively few paragraphs over a half-page long?

- Appropriate use of verb tense
- A coherent flow, linking new information to known or previously given information?
- Correct grammar, punctuation, and spelling

14. Appendices

- Included as necessary to provide detailed additional information that is not essential in the body of the report but that would be of interest for specialist readers?
- Informatively titled and labelled as Appendix A, Appendix B, etc., according to the order in which they are mentioned in the text?

15. Formatting of the report

- Use typography, lists, and informative headings to make the structure of the report clear, the appearance attractive, and the information accessible?

16. Information from primary and secondary sources used in the report

- Used effectively
- Formatted so that wording taken directly from the original source appears in quotation marks or inset block format
- Whether quoted or paraphrased, accompanied by an in-text citation or reference number acknowledging the source according to a standard format (e.g., APA or IEEE) and directing the reader to more complete publication information included.

Short Question and Answers

1. Write a brief note on Footnotes in the context of a Research Report.

Ans :

Footnotes are notations (notes) at the end of the page in a printed document. They are presented in a smaller font size traits the dominant text. They are indicated with superscript numbers or various symbols. When many footnotes are present in a text, numbers are used to indicate them and when few footnotes are present, symbols are used to indicate them. The asterisk (*) symbol is most commonly used for footnotes.

The use of footnotes in the research report helps in servicing the two important objectives first the recognition of the source of materials used in quotations are made so that the reader can easily cross verily any references. Further footnotes provide supplemental values to the report.

The researcher needs to take into consideration the following points with reference to the footnotes,

(i) A footnotes may be separated from the textual material by a line, which is spaced half an inch and a line about one and a half inches long from the textual material.

(ii) All the footnotes must be numbered in a sequential order usually starting with 1 in each chapter separately.

(iii) In case of statistical tables or other numerical material, an asterisk symbol (*) or a similar type of a symbol may be used in order to avoid confusion.

2. Reports.

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, planning future course of action and taking effective decisions. Reports acts as a tool for providing feedback to the managers related to various aspects of the organization.

3. Define Interpretation.

Ans :

Meaning

Interpretation refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study. In fact, it is a search for broader meaning of research findings. The task of interpretation has two major aspects viz.,

(i) The effort to establish continuity in research through linking the results of a given study with those of another, and

(ii) The establishment of some explanatory concepts. Interpretation is concerned with relationships within the collected data and the extension of study beyond the collected data as well.

Thus, interpretation is the device through which the factors that seem to explain what has been observed by researcher in the course of the study can be better understood and it also provides a theoretical conception which can serve as a guide for further researches.

4. Define Statistical Fallacies.

Ans :

Interpretation of data, is a very difficult task and requires a high degree of care, objectivity, skill and judgement. In the absence of these things, it is likely that the data may be misused. In fact, experience shows that the largest number of mistakes are committed knowingly or unknowingly while interpreting statistical data which may lead to misinterpretation of data by most of the readers.

Statistical fallacies may arise at any stage – in the collection, presentation, analysis and interpretation of data. The following are some of the

(i) Specific examples illustrating how statistics can be misinterpreted,

(ii) Sources of errors leading to false generalizations,

(iii) Examples how fallacies arise in using statistical data and statistical methods.

5. What is Report Writing ? Explain the objectives of report writing.

Ans :

Report writing is a conscious, rational and systematic effort. It is both an art as well as science. Report-writing requires conceptual and communication skills and a scientific approach to investigation, analysis and presentation. Managers should possess right abilities and attitudes for creating effective reports.

Objectives

Reports are written for various purposes such as,

(i) For reviewing performance

(ii) Keep a check on a continuing activity

(iii) Survey the market

(iv) Future planning needs of the organization

(v) Submit standardized information etc.

6. What is Bibliography?

Ans :

It refers to the list of references which are related to a topic or subject. It includes all the information which is found in the first footnote and it is related to the research work. It is present at the end of the main body of a report. Bibliography differs in its functions from that of the footnotes.

The main function of the bibliography is to list in an alphabetical order all the references which are being used by the researcher. But the foot notes only provides the identification details of a work.

7. Bias

Ans :

Bias is very common in statistical work and often leads to false conclusions. For example, if a researcher wants to show that the wages at a particular company are very low he may select the sample in such as manner as to exclude high paid workers. Similarly, businessmen may use statistics to prove the superiority of their own business over others.

8. Discuss the method of generalization

Ans :

Normally, two methods are used for generalization viz.

1. Logical Method

(i) The Method of Agreement

(ii) The Method of Difference

(iii) Joint Method of Agreement and Difference

(iv) The Method of Residues

(v) The Method of Concomitant Variation

2. Statistical method

(i) Collection of Data

(ii) Presentation of Data

(iii) Analysis of Data

(iv) Interpretation of Data

9. Routine Reports.

Ans :

Routine reports consists of forms in which blanks need to be filled in or multiple choice statements are to be ticked. Report forms are prepared efficiently because it gives complete information to management. The different types of routine reports are,

(a) Progress Report

Progress report provides information regarding progress of a project during completion of the project like manufacture of products or implementation of a scheme, construction of a building etc.

(b) Inspection Report

After the completion of inspection, inspection report is submitted. Inspection is necessary for determining irregularities or deviation from standard practice in the daily activities of the company.

Example: Internal audit reports.

Inspection of machinery, buildings and property are conducted at periodic intervals. Audits and inspection of work are done to know whether the work is carried out properly or not.

(c) Performance Appraisal

Performance appraisal is prepared by filling a form periodically. It helps in assessing the performance of an employee. Supervisor job is to fill an assessment report regarding the working of the subordinates. Performance appraisal thus help superiors to assess the performance and to provide feedback on their performance. Performance appraisal reports helps in taking decisions related to promotions and other employee benefits.

10. Survey Report.

Ans :

This type of report is prepared when an organization launches a new product in the market, introduce a new service or any major changes which have its impact on company's customer.

The factors to be examined totally depends on the purpose of the survey. The purpose of conducting this report may be due to the factors like suitability of a site for a factory, to evaluate the feasibility and financial viability of a proposal, to survey the market, to estimate damage etc.

Choose the Correct Answers

1. Report is a _____ and compact written document giving updated information about a specific problem. [b]

   (a) Research (b) Complete
   (c) Report (d) Impact

2. Compact and self-explanatory document over a _____ period. [a]

   (a) Long (b) Short
   (c) High (d) Low

3. A report serves as a permanent record relating to certain _____ matter. [c]

   (a) Agriculture (b) Statistical
   (c) Business (d) Research

4. A technical _____ has a number of clearly defined sections. [b]

   (a) Research (b) Report
   (c) Business (d) Agriculture

5. _____ is a list of references relating to a topic subject. [a]

   (a) Bibliography (b) Classification
   (c) Cronology (d) Biography

6. _____ presents specialized terms in alphabetical order. [a]

   (a) Glossory (b) Figures
   (c) Tables (d) Body

7. Use of index numbers are also lead to _____ conclusions. [b]

   (a) Correct (b) Wrong
   (c) Negative (d) Positive

8. Coefficient of association may also give fallacious _____. [a]

   (a) Conclusions (b) Decisions
   (c) Orders (d) Reports

9. In certain cases conclusions based on averages lead to _____ inferences. [d]

   (a) Correct (b) Negative
   (c) Positive (d) Wrong

10. The work of interpretation of statistical data is of a complicated _____. [a]

    (a) Nature (b) Climate
    (c) Culture (d) None

Fill in the blanks

1. A report has only one function to perform. It must _____
2. A treatise on a single subject is called a _____
3. The contents of a report can broadly be classified into _____ parts.
4. An abstract is a _____
5. An Index may be either _____ index or _____ index.
6. A _____ contains the sources of references cited and other relevant works consulted.
7. The list of special terms and phrases used is given in the form a _____
8. A _____ report is a formal statement of the research process and its results.
9. A report serves as a means for judging the _____ of the completed research project.
10. A presentation of the salient findings with _____.

ANSWERS

1. Inform
2. Monograph
3. Three
4. Synopsis
5. Subject, author
6. Bibliography
7. Glossary
8. Research
9. Quality
10. Tables

One Mark Answers

1. Define Interpretation.

Ans :

Interpretation refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study. In fact, it is a search for broader meaning of research findings.

2. Statistical Fallacies.

Ans :

Interpretation of data, is a very difficult task and requires a high degree of care, objectivity, skill and judgement. In the absence of these things, it is likely that the data may be misused. In fact, experience shows that the largest number of mistakes are committed knowingly or unknowingly while interpreting statistical data which may lead to misinterpretation of data by most of the readers.

3. Routine Reports.

Ans :

Routine reports consists of forms in which blanks need to be filled in or multiple choice statements are to be ticked. Report forms are prepared efficiently because it gives complete information to management.

4. Periodical Report

Ans :

Periodical reports are made at regular intervals and provides information about the day-to-day work. These reports are usually prepared by filling in a form.

5. Bibliography

Ans :

It refers to the list of references which are related to a topic or subject. It includes all the information which is found in the first footnote and it is related to the research work. It is present at the end of the main body of a report. Bibliography differs in its functions from that of the footnotes.

UNIT III

Statistical Estimation, Hypothesis Testing and Sampling of Variables:

Statistical Estimation: Concepts of Population, Sample and Sampling Distribution- Parameters and Statistics - Central limit theorem - Concept of Standard Error - Confidential Limits - Estimation of Population Parameters - Properties of a Good Estimator - Point and Interval Estimation - -

Hypothesis Testing: Formulation of Hypotheses: Types of Hypotheses - Methods of testing Hypotheses - Type I and Type II errors - One tail and two tail tests (Theory only). Sampling of Attributes: Estimation and testing of Number and Proportions of Success - Difference between two proportions (including problems).

Sampling of Variables: Large Samples: Difference between large and small samples - Estimating population mean - Testing: Significance of Mean - Significance of the difference between means of two samples - Significance of the difference between the standard deviations of two samples. (Including problems) Small Samples: 't' test - Fixing fiducial limits to population mean – Testing: Significance of the mean - Significance of the difference between two independent means – Significance of the difference between two dependent means (including problems).

3.1 STATISTICAL ESTIMATION

3.1.1 Concepts of Population, Sample and Sampling Distribution - Parameters and Statistics

Q1. Define the following terms :

(a) Population and sample

(b) Sampling Distribution

(c) Statistics and Parameters

Ans :

(a) Population and Sample

Definition

- Population (or universe) is the aggregate or totality of statistical data forming a subject of investigation.

  For example

  (i) The population of the heights of Indians,

  (ii) The population of Nationalized Banks in India, etc.

- The number of observations in the population is defined to be the size of the population. It may be finite or infinite. Size of population is denoted by N.

  Examples, if there are 600 students in the school that we classified according to block type, we say that we have a population of size 600. The numbers on the cards in a deck, the heights of residents in a certain city, and the lengths of fish in a particular lake are all examples of population with finite size.

  The observations obtained by measuring the atmospheric pressure every day from the past on into the future, or all measurements on the depth of a lake from any conceivable. position, are examples of population whose sizes are infinite.

  In the field of statistical inference the statistician is interested in arriving at conclusion concerning a population when it is impossible or impractical to observe the entire set of observations that make up the population.

  For example, in attempting to determine the average length of life of a certain brand c light bulb, it would be impossible to test all such bulbs if we are to have any left to sell.

  Therefore, we must depend on a subset of observations from the population to help us make inferences concerning the same population. This brings us to consider the notion of sampling.

(b) Sampling Distribution

Sampling theory is the study of relationships between a population and samples drawn from the population, and it is applicable to random

samples only. We need to upto maximum information about the population with the help of samples, i.e., to determine the true value of the population parameters.

(c) Statistics and Parameters

A statistic is a characteristic of a sample, whereas a parameter is a characteristic of a population. Thus, when we work out certain measures such as mean, median, mode or the like ones from samples, then they are called statistic(s) for they describe the characteristics of a sample. But when such measures describe the characteristics of a population, they are known as parameters.

3.2 Central Limit Theorem

Q2. Explain briefly about Central Limit Theorem with an example.

Ans : **(Imp.)**

The Central Limit Theorem is the sampling distribution of the sampling means approaches a normal distribution as the sample size gets larger, no matter what the shape of the data distribution. An essential component of the Central Limit Theorem is the average of sample means will be the population mean.

Similarly, if you find the average of all of the standard deviations in your sample, you will find the actual standard deviation for your population.

i) Mean of sample is same as the mean of the population.

ii) The standard deviation of the sample is equal to the standard deviation of the population divided by the square root of the sample size.

Central limit theorem is applicable for a sufficiently large sample sizes (n = 30). The formula for central limit theorem can be stated as follows:

$$\mu_{\bar{x}} = \mu$$

and

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Where,

μ = Population mean

σ = Population standard deviation

$\mu_{\bar{x}}$ = Sample mean

$\sigma_{\bar{x}}$ = Sample standard deviation

n = Sample size

Example

The record of weights of the male population follows the normal distribution. Its mean and standard deviations are 70 kg and 15 kg respectively. If a researcher considers the records of 50 males, then what would be the mean and standard deviation of the chosen sample?

Sol :

Mean of the population μ = 70 kg

Standard deviation of the population = 15 kg

sample size n = 50

Mean of the sample is given by:

$\mu_{\bar{x}}$ = 70 kg

Standard deviation of the sample is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{15}{\sqrt{50}}$$

$$= \frac{15}{7.07}$$

$\sigma_{\bar{x}}$ = 2.122 = 2.1 kg (approx)

3.3 Concept of Standard Error - Confidential Limits

Q3. Describe briefly about the concept of standard error and confidential limits.

Ans : **(Imp.)**

i) The standard deviation of sampling distribution of a statistic is known as its standard error (S.E) and is considered the key to sampling theory.

ii) The utility of the concept of standard error in statistical induction arises on account of the following reasons:

iii) The standard error helps in testing whether the difference between observed and expected frequencies could arise due to chance.

iv) The criterion usually adopted is that if a difference is less than 3 times the S.E., the difference is supposed to exist as a matter of chance and if the difference is equal to or more than 3 times the S.E., chance fails to account for it, and we conclude the difference as significant difference.

v) This criterion is based on the fact that at $\bar{X} \pm 3$ (S.E.) the normal curve covers an area of 99.73 per cent. Sometimes the criterion of 2 S.E. is also used in place of 3 S.E.

vi) Thus the standard error is an important measure in significance tests or in examining hypotheses.

vii) If the estimated parameter differs from the calculated statistic by more than 1.96 times the S.E., the difference is taken as significant at 5 per cent level of significance.

viii) This, in other words, means that the difference is outside the limits i.e., it lies in the 5 per cent area (2.5 per cent on both sides) outside the 95 per cent area of the sampling distribution.

ix) Hence we can say with 95 per cent confidence that the said difference is not due to fluctuations of sampling.

x) In such a situation our hypothesis that there is no difference is rejected at 5 per cent level of significance. But if the difference is less than 1.96 times the S.E., then it is considered not significant at 5 per cent level and we can say with 95 per cent confidence that it is because of the fluctuations of sampling. In such a situation our null hypothesis stands true. 1.96 is the critical value at 5 per cent level.

xi) The product of the critical value at a certain level of significance and the S.E. is often described as 'Sampling Error' at that particular level of significance.

xii) Then can test the difference at certain other levels of significance as well depending upon our requirement.

xiii) The following table gives some idea about the criteria at various levels for judging the significance of the difference between observed and expected values:

Criteria for judging Significance at Various Important Levels

Significance level	Confidence level	Critical value	Sampling error	Confidence limits	Difference Significance if	Difference Insignificant if
5.0%	95.0%	1.96	1.96σ	$\pm 1.96\sigma$	$\geq 1.96\sigma$	$\leq 1.96\sigma$
1.0%	99.0%	2..5758	$2..5758\sigma$	$\pm 2.5758\sigma$	$\geq 2.5758\sigma$	$\leq 2.5758\sigma$
2.7%	99.73%	3	3σ	$\pm 3\sigma$	$\geq 3\sigma$	$\leq 3\sigma$
4.55%	95.45%	2	2σ	$\pm 2\sigma$	$\geq 2\sigma$	$\leq 2\sigma$

σ = Standard Error

The standard error gives an idea about the reliability and precision of a sample. The smaller the S.E., the greater the uniformity of sampling distribution and hence, greater is the reliability of sample. Conversely, the greater the S.E., the greater the difference between observed and expected frequencies. In such a situation the unreliability of the sample is greater. The size of S.E., depends upon the sample size to a great extent and it varies inversely with the size of the sample. If double reliability is required i.e., reducing S.E. to 1/2 of its existing magnitude, the sample size should be increased four-fold.

The standard error enables us to specify the limits within which the parameters of the population are expected to lie with a specified degree of confidence. Such an interval is usually known as confidence interval. The following table gives the percentage of samples having their mean values within a range of population mean (μ) ± S.E.

Range	Per cent Values
$\mu \pm 1$ S.E.	68.27%
$\mu \pm 2$ S.E.	95.45%
$\mu \pm 3$ S.E.	99.73%
$\mu \pm 1.96$ S.E	95.00%
$\mu \pm 2.5758$ S.E	99.00%

3.4 Estimation of Population Parameters

3.4.1 Types of Estimation

3.4.1.1 Point and Interval Estimation

Q4. Define the term Estimation. What are the different types of Estimations?

(OR)

Define :

(a) Point Estimation

(b) Interval Estimation

Ans : **(Imp.)**

Meaning

Quantities appearing in distributions, such as p in the binomial distribution and μ and σ in the normal distribution are called parameters.

(i) Estimate

An estimate is a statement made to find an unknown population parameter.

(ii) Estimator

The procedure or rule to determine an unknown population parameter is called an estimator. For instance, sample mean is an estimator of population mean because sample mean is a method of determining the population mean. Remember that an estimator must be a statistic and it must depend only on the sample and not on the parameter to the estimated. So an estimator is a statistic which for all practical purposes, can be used in place of unknown parameter of the population.

(a) Point Estimation: A point estimate is a single number which is used as an estimate of the unknown population parameter. The procedure in point estimation is to select a random sample of n observations, $x_1, x_2 ... x_3$ from a population $f(x, \theta^*)$ and then to use some preconceived method to arrive from these observations at a number say $\hat{\theta}$ (read theta hat) which we accept as an estimator of θ. The estimator θ is a single point on the real scale and thus the name point estimation, $\hat{\theta}$ depends on the random variables that generate the sample and hence, it too is a random variable with its own sampling distribution.

(b) Interval Estimation: An interval estimate of a population parameter is a statement of two values between which it is estimated that the parameter lies. An interval estimate would always be specified by two values, i.e., the lower one and the upper one. In more technical terms, interval, called the confidence interval, whose end points L and U with $L < U$ are functions of the observed random variables such that the probability that the inequality $L < \theta < U$ is satisfied in terms of pre-determined number.

3.4.2 Properties of a Good Estimator

Q5. Explain the properties of a Good Estimator.

(OR)

State the properties of a Good Estimator.

Ans : **(Imp.)**

A good estimator, as common sense dictates, is close to the parameter being estimated. Its quality is to be evaluated in terms of the following properties

(i) Unbiasedness

An estimator is said to be unbiased it its expected value is identical with the population parameter being estimated. That is $\hat{\theta}$ is an unbiased estimate of θ, then we must have $E(\hat{\theta}) = \theta$. many estimators are Asymptomatically unbiased in the sense that the biases reduce to practically insignificant values zero when n becomes sufficiently large.

It should be noted that bias in estimation is not necessarily undersiable. It may turn out be an asset in some situations.

(ii) Consistency

If an estimator, say $\hat{\theta}$, approaches the parameter θ closer and closer as the sample size n increases, $\hat{\theta}$ is said to be a consistent estimator of θ. Stating somewhat more regorously, the estimator $\hat{\theta}$ is said said to be a consistent estimator of θ if, as n approaches infinity, the probability approaches 1 that $\hat{\theta}$ will differ from the parameter θ by not more than an arbitrary small constant.

The sample mean is an unbiased estimator of μ no matter what form the population distribution assumes, while the sample median is an unbiased estimate of μ only if population distribution is symmetrical. The sample mean is better that the sample median as an estimated of μ in terms of both unbiasedness and consistency.

(iii) Efficiency

The concept of efficiency refers to the sampling variability of an estimator. If two competing estimators are both unbiased, the one with the smaller variance (for a given sample size) is said to be relatively more efficient. Stated in a somewhat different language, estimator $\hat{\theta}_1$ is said to be more efficient than another estimator $\hat{\theta}_2$ for θ if the variance of the first is less than the variance of the second. The smaller the variance of the estimator, the more concentrated is the distribution of the estimator around the parameter being estimated and therefore, the better this estimator is.

(iv) Sufficiency

An estimator is said to be sufficient if it conveys as much information as is possible about the parameter which is contained in the sample. The significance of sufficiency lies in the fact if a sufficient estimator exists, it is absolutely unnecessary to consider any other estimator a sufficient estimator ensures that all information a sample can furnish with respect to the estimation of a parameters is being utilized.

Many methods have been devised for estimating parameters that may provide estimators satisfying these properties. The two important methods are the least square method and the method of maximum likehood.

3.5 HYPOTHESIS TESTING

3.5.1 Formulation of Hypothesis

Q6. What is a Hypothesis? State the characteristics and importance of hypothesis.

Ans :

Definitions

Hypothesis is defined in different ways by different people, some of them are as follows,

i) **According to James E. Greighton,** "It is a tentative supposition or provisional guess which seems to explain the situation under observation".

ii) **According to Bruce W. Titckman,** "A hypothesis then could be defined as an expectation about events based on generalization of the assumed relationship between variables".

Examples

(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.

Characteristics

Some of the key features of a hypothesis are as follows,

1. Hypothesis is conceptual in nature as few conceptual elements are included in its framework.
2. It is basically a verbal statement in a declarative type, which has a verbal expression of ideas and concepts which are ready for empirical inspection.
3. It holds empirical referent which highlights the tentative relationship between two (or) more variables.
4. It is future oriented in nature and is associated with the future verification and does not take into consideration the historical facts and informations.
5. It is an important component of a scientific research and all the research activities are formulated for its verification.

Importance

The importance of hypothesis are as follows,

1. Hypothesis basically acts as the eyes of investigator by supporting him in finding out answers for tentatively following generalization.
2. Hypothesis usually emphasizes upon research. In absence of research hypothesis there would be a more waiting.
3. It acts as an important link between theory and the investigation.
4. A well planned hypothesis usually provides a clear and particular objectives to the research worker and facilitates him by offering a base for sample selection and research method for achieving the set objectives.
5. Hypothesis acts as the key function of clubbing together all the associated facts and information into one.
6. Hypothesis helps in avoiding a blind search and random collection of data which might later on tends to be unnecessary for the problem which is being studied.
7. Hypothesis acts as one of the powerful tool which helps in showing right path for the research work.

3.5.2 Types of Hypotheses

Q7. Explain the various types of hypothesis with examples.

Ans :

Hypothesis can be studied under the following four types,

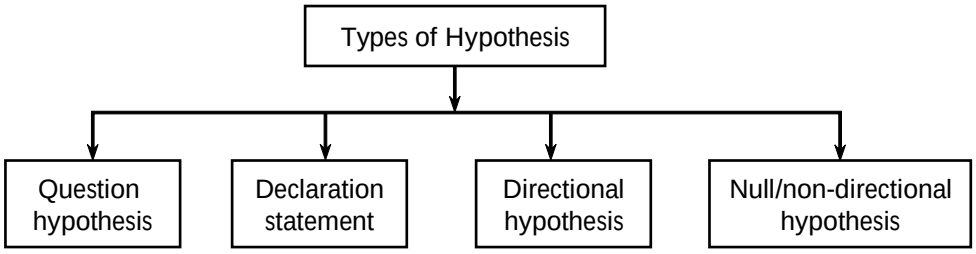


Fig.: Types of Hypothesis

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.

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."

3.5.3 Methods of Testing Hypothesis

Q8. Explain various methods of Testing Hypothesis.

Ans : **(Imp.)**

Types of Hypothesis Testing

There are several types of hypothesis testing, and they are used based on the data provided. Depending on the sample size and the data given, we choose among different hypothesis testing methodologies. Here starts the use of hypothesis testing tools in research methodology.

i) Normality: This type of testing is used for normal distribution in a population sample. If the data points are grouped around the mean, the probability of them being above or below the mean is equally likely. Its shape resembles a bell curve that is equally distributed on either side of the mean.

ii) T-test: This test is used when the sample size in a normally distributed population is comparatively small, and the standard deviation is unknown. Usually, if the sample size drops below 30, we use a T-test to find the confidence intervals of the population.

iii) Chi-Square Test: The Chi-Square test is used to test the population variance against the known or assumed value of the population variance. It is also a better choice to test the goodness of fit of a distribution of data. The two most common Chi-Square tests are the Chi-Square test of independence and the chi-square test of variance.

iv) ANOVA: Analysis of Variance or ANOVA compares the data sets of two different populations or samples. It is similar in its use to the t-test or the Z-test, but it allows us to compare more than two sample means. ANOVA allows us to test the significance between an independent variable and a dependent variable, namely X and Y, respectively.

v) Z-test-: It is a statistical measure to test that the means of two population samples are different when their variance is known. For a Z-test, the population is assumed to be normally distributed. A z-test is better suited in the case of large sample sizes greater than 30. This is due to the central limit theorem that as the sample size increases, the samples are considered to be distributed normally.

3.5.4 Type I and Type II Errors

Q9. Explain briefly about Type I and Type II Errors

Ans :

There are basically two types of errors we make in the context of testing of Hypothesis. These are called as Type-I error and the Type-II error. In type-I error, we may reject Null hypothesis when Null hypothesis is true. Type-II error is when we accept Null hypothesis when the Null hypothesis is not true. In other words, Type-I error means rejection of hypothesis which should have been accepted and Type-II error means accepting the hypothesis which should have been rejected. Type-I error is denoted by alpha known as alpha error, also called the level of significance of test and Type-II error is denoted by beta known as beta error

Nature	Accept Null hypothesis	Reject Null hypothesis
Null hypothesis (true)	Correct decision	Type-I error (alpha error)
Null hypothesis (false)	Type-II error (beta error)	Correct decision

The probability of Type-I error is usually determined in advance and is understood as the level of significance of testing the hypothesis. If Type-I error is fixed at 5%, it means that there are about 5 chance in100 that we will reject Null hypothesis when Null hypothesis is true. We can control Type-I error just by fixing at a lower level. For instance, if we fix it at 1%, we will say that the maximum probability of committing Type-I error would only be 0.01.

But with the fixed sample size, when we try to reduce Type-I error, the probability of committing Type-II error increases. Both types of errors cannot be reduced simultaneously. There is trade off between two types of errors which means that the probability of making one type error can only be reduced if we are willing to increase the probability of making the other type of error. One must set a very high level for Type-I error in one's testing technique of a given hypothesis. Hence, in the testing of hypothesis, one WWW must make all possible efforts to strike an adequate balance between Type-I and Type-II errors.

3.5.5 One Tail and Two Tail Tests

Q10. Explain briefly about One Tailed and Two Tailed Tests

Ans : **(Imp.)**

(i) One tailed test

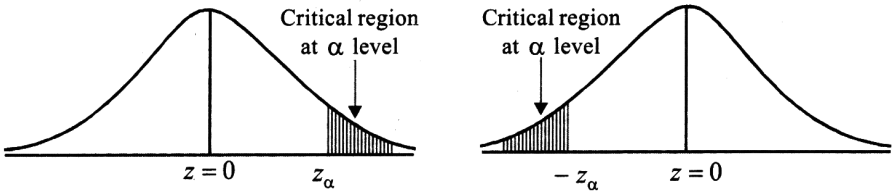
If we have to test whether the population mean μ has a specified value μ_0, then the Null Hypothesis is $H_0 : \mu = \mu_0$ and the Alternative Hypothesis may be

(i) $H_1 : \mu \neq \mu_0$ (i.e., $\mu > \mu_0$ or $\mu < \mu_0$) or

(ii) $H_1 : \mu > \mu_0$ or

(iii) $H_1 : \mu < \mu_0$.

The Alternative Hypothesis in (i) is known as a two-tailed (i.e., both right and left tail) alternatives and the alternative hypothesis n (ii) and (iii) are known as right-tailed and left-tailed alternatives respectively.



If the Alternative Hypothesis H_1 in a test of a statistical hypothesis be one-tailed i.e., either right- tailed or left-tailed but not both), then the test is called a one-tailed test.

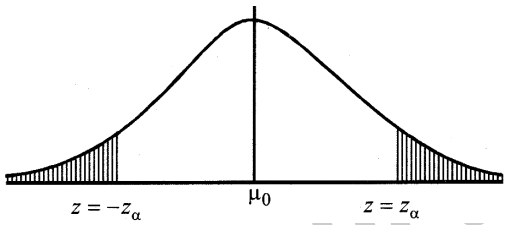
For example, to test whether the population mean $\mu = \mu_0$, we have $H_0 : \mu = \mu_0$ against the alternative hypothesis H_x given by

(a) $H_1 : \mu > \mu_0$ (right tailed) or

(b) $H_1 : \mu > \mu_0$ (left-tailed) and the corresponding test is a single-tailed or one-tailed or one sided. In the right-tail test $H_x : \mu > \mu_0$, the critical region (or rejection region) $z > z_\alpha$ lies entirely in the right tail of the sampling distribution of sample mean x with area equal to the level of significance α (see fig). Similarly, in the left - tailed test ($H_1 : \mu < \mu_0$), the critical region $z < -z_\alpha$ lies entirely in the left tail of the sampling distribution of the sample mean $\bar{x}$ with area equal to the level of significance a (see fig).

(ii) Two-Tailed Test

Suppose we want to test the Null Hypothesis $H_0 : \mu = \mu_0$ against the Alternative Hypothesis $H_1 : \mu \neq \mu_0$.



Since H_1 is two-tailed alternative hypothesis, the critical region under the curve is equally distributed on both sides of the mean.

Thus, the critical area under the right - tail = The critical area under the left - tail

= Half of the total area

$= \frac{1}{2}$ probability of rejection

$= \frac{\alpha}{2}$

with critical statistic $Z_{\alpha/2}$, where α is the level of significance.

The critical region is then, $z \leq -z_{\alpha/2}$ or $z_{\alpha/2} \leq z$

Critical values of x for both two-tailed and one-tailed tests at 1%, 5% and 10% level of significance are given in the following table :

Critical values of Z

Level of significance α	**1%**	**5%**	**10%**
Critical values for two-tailed test	$\|z_\alpha\| = 2.58$	$\|z_\alpha\| = 1.96$	$\|z_\alpha\| = 1.645$
Critical values for Right-tailed test	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Critical values for Left-tailed test	$z_\alpha = -2.33$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

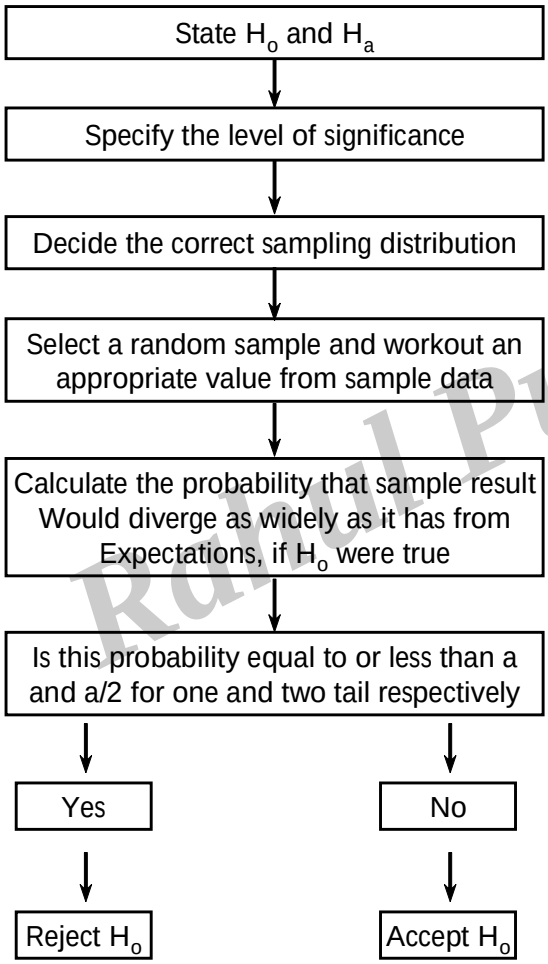
Applying one-tailed or two-tailed test for a particular problem depends entirely on the nature of the Alternative Hypothesis. If the alternative test is two-tailed we apply two-tailed test and if Alternative Hypothesis is one-tailed, we apply one-tailed test.

Q11. Explain the procedure and rules of testing hypothesis.

Ans :

Procedure of Hypothesis Testing

Following are the steps involved in testing of hypothesis,



1. Making a Formal Statement

This step consists of making a formal statement of the null hypothesis (H_0) and also of alternatively hypothesis (H_a). This means that the hypothesis should be clearly stated, considering the nature of the research problem.

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.

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 α value in case of one tailed test and $\frac{\alpha}{2}$ value in case of two tailed test, then reject H_0 and accept H_a and vice versa.

Rules

Rule 1 : Null Hypothesis

To test the null hypothesis that the population has a specified mean μ_0 then null hypothesis is,

$H_0: \mu = \mu_0$

Rule 2 : Alternative Hypothesis

Set up the alternative hypothesis = H_1 which helps us to decide whether we have to use,

(i) $H_1 : \mu \neq p_0$ (Two tailed alternative)

(ii) $H_1 : \mu > \mu_0$ (Right tailed alternative)

(iii) $H_1 : \mu < \mu_0$ (Left tailed alternative).

Rule 3 : Level of Significance

Choose an appropriate level of significance (α). The level of significance usually employed are 5% and 1%.

Rule 4 : Test Statistic

Evaluate the value of Z, i.e., the test statistic under the null hypothesis.

$$Z = \frac{t - E(t)}{S.E(t)} = \frac{t - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Rule 5: Decision

The computed value of the test statistic Z is compared with the critical value Z_α at given level of significance (α). Then we draw following two conditions,

If $|Z| < Z_\alpha$

If the computed value of Z is less than the critical value Z_α, we say that it is not significant. Therefore, we "accept the null hypothesis".

If $|Z| > Z_\alpha$

If the computed value Z is greater than the critical value Z_α, we say that it is significant and hence, the "null hypothesis is rejected" at the given level of significance (α).

3.6 SAMPLING OF ATTRIBUTES

Q12. What do you understand by Sampling of Attributes?

Ans : **(Imp.)**

Meaning

The sampling of attributes may, therefore, be regarded as the drawing of samples from a population whose members possess the attribute A or not A.

For example, in the study of attribute 'Literacy' a sample may be taken and people classified as literates and illiterates. With such data the binomial type of problem may be formed. The selection of an individual on sampling may be called 'event', the appearance of attribute A may be taken as 'success' and its non-appearance as 'failure'. Thus, if out of 1,000 people selected for the samples, 100 are found literates, and 900 illiterates, we would say that the sample consists of 1000 units out of which 100 are successes and 900 failures. The probability of success or p = 100 / 1,000 or 0.1 and the probability of failure or q = 900 / 1,000 = 0.9 so that p + q = 0.1 + 0.9 = 1.

The various tests of significance for attributes are discussed under the following heads :

(i) Test for Number of Successes

(ii) Tests for proportion of Successes

(iii) Tests for difference between Proportion

3.6.1 Estimation and Testing for Number of Successes

Q13. How to estimate the number of successes for attributes.

Ans :

The sampling distribution of the number of successes follows a binomial probability distribution. Hence its standard error is given by the formula :

S.E of no. of successes = $\sqrt{npq}$

where

n = size of sample

p = probability of success in each trial

q = (1 – p), i.e., probability of failure.

PROBLEMS

1. A coin was tossed 400 times and the head turned up 215 times. Test the hypothesis that the coin is unbiased.

Sol :

Let us take the hypothesis that the coin is unbiased. On the basis of this hypothesis the probability of getting head or tail would be equal, i,e., ½ . Hence in 400 throws of a coin we should expected 200 heads and 200 tails.

Observed number of heads = 216

Difference between observed number of heads and expected number of heads = 216 – 200 = 16

S.E. of no. of heads = $\sqrt{npq}$

$$n = 400\frac{1}{2} \quad P = q = \frac{1}{2}$$

$$\therefore \quad S.E. = \sqrt{400 \times \frac{1}{2} \times \frac{1}{2}} = 10$$

$$\frac{\text{Difference}}{\text{S.E.}} = \frac{16}{10} = 1.6$$

The difference between observed and expected number of heads is less than 1.96 SE (5% level of significance). Hence our hypothesis is true and we conclude that the coin is unbiased.

2. In 324 throws of a six-faced dice, odd points appeared 180 times. Would you say that the dice is fair at 5 per cent level of significance ?

Sol :

Let us take the hypothesis that the dice is fair. In a fair dice we would expect 162 odd points in 324 throws

$$S.E. = \sqrt{npq}$$

$$n = 324 \quad P = \frac{1}{2} \quad q = \frac{1}{2}$$

$$= \sqrt{324 \times \frac{1}{2} \times \frac{1}{2}} = 9$$

$$\frac{\text{Difference}}{\text{S.E}} = \frac{180 - 162}{9} = \frac{18}{9} = 2$$

Since the difference is more than 1.96 at 5 per cent level of significance, the hypothesis rejected. Hence we cannot say that the dice is fair at 5 per cent level of significance.

3.6.2 Estimation and Testing for Proportion of Successes

Q14. How to estimate the number of Proportion for attributes.

Ans :

Instead of recording the number of successes in each sample, we night record the proportion successes, that is $\frac{1}{n}$ th of the number of successes in each sample. As this would amount to dividing all figures of the record by n, the mean proportion of successes must be p and the standard deviation of the proportion of successes $\sqrt{\frac{pq}{n}}$.

Thus we have the following formula :

$$S.E.\ p = \sqrt{\frac{pq}{n}}$$

PROBLEMS

3. 500 apples are taken at random from a large basket and 50 are found to be bad. Estimate the proportion of bad apples in the basket and assign limits within which the percentage most probably lies.

Sol : **(Imp.)**

The proportion of bad apples in the given sample $= \frac{50}{500} = 0.1$

Hence $p = 0.1$ and $q = 0.9$

$$\text{S.E.} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.1 \times 0.9}{500}} = \sqrt{\frac{0.09}{500}} = 0.013$$

The limits within which percentage of bad apples lies

$$= \left[p \pm 3\sqrt{\frac{pq}{n}} \right] \times 100$$

$= 0.1 \pm 3(0.013)] \times 100$

$= 0.1 \pm (0.039)] \times 100$

$= 4$ and 3.8

Thus the percentage of bad apples in the consignment almost certainly lies between 4 and 3.8.

3.6.3 Test for Difference between Two Proportions

Q15. How to estimate the Test for difference between Two Proportions.

Ans :

If two samples are drawn from different populations, we may be interested in finding out whether the difference between the proportion of successes is significant or not. In such a case we take the hypothesis that the difference between p_1, i.e., the proportion of successes in one sample, and p_2 i.e., the proportion of successes in another sample, is due to fluctuations of random sampling.

The standard error of the difference between proportions is calculated by applying the following formula :

$$\text{S.E.}\ (p_1 - p_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where p = the pooled estimate of the actual proportion in the population.

The value of p is obtained as follows :

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \text{ (or) } p = \frac{x_1 + x_2}{n_1 + n_2}$$

where x_1 and x_2 stand for the number of occurrences in the two samples of sizes n_1 and n_2 respectively.

If $\frac{p_1 - p_2}{\text{S.E.}}$ is less than 1.96 S.E. (5% level of significance), the difference is regarded as due to random sampling variation, i.e., as not significant.

PROBLEMS

4. In a random sample of 1,000 persons from town A, 400 are found to be consumers of wheat. In a sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B, so far as the proportion of wheat consumers is concerned ?

Sol : **(Imp.)**

Let us set up the hypothesis that the two towns do not differ so far as proportion of wheat consumers is concerned, i.e., $H_0 : p_1 = p_2$ against $H_a = p_1 \neq p_2$.

Computing the standard error of the difference of proportions :

$$\text{S.E. } (p_1 - p_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$n_1 = 1000,\ p_1 = \frac{400}{1000} = 0.4;\ n_2 = 800;\ p_2 = \frac{400}{800} = 0.5$$

$$p = \frac{(100 \times 0.4) + (800 \times 0.5)}{1000 + 800}$$

$$= \frac{400 + 400}{1800}$$

$$= \frac{800}{1800} = \frac{4}{9}$$

$$\text{or simply } p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{400 + 400}{1000 + 800} = \frac{800}{1800} = \frac{4}{9}$$

$$q_1 = 1 - \frac{4}{9} = \frac{5}{9}$$

$$\text{S.E. } (p_1 - p_2) = \sqrt{\frac{4}{9} \times \frac{5}{4}\left(\frac{1}{1000} + \frac{1}{800}\right)}$$

$$= \sqrt{\frac{20}{81} \times \frac{9}{4000}} = 0.024$$

$$p_1 - p_2 = 0.4 - 0.5 = -0.1$$

$$\frac{\text{Difference}}{\text{S.E.}} = \frac{0.1}{0.024} = 4.17.$$

Since the difference is more than 2.58 S.E. (1% level of significance). It could not have arisen due to fluctuations of sampling. Hence the data reveal a significant difference between town A and town B so far as the proportion of wheat consumers is concerned.

5. In a simple random sample of 600 men taken from a big city 400 are found to be smokers. In another simple random sample of 900 men taken from another city 450 are smokers do the data indicate that there is a significant difference in the habit of smoking in the two cities ?

Sol :

Let us take the hypothesis that there is no significant difference in the habit of smoking in the two cities.

$$\text{S.E. } (p_1 - p_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$p_1 = \frac{400}{600} = 0.667$$

$$p_2 = \frac{400}{900} = 0.44$$

$$p = \frac{x_1 + x_2}{n_1 + n_2}, \; q = 1 - p$$

$$n_1 = 600, \; n_2 = 900, \; x_1 = 400, \; x_2 = 450$$

$$p = \frac{400 + 450}{600 + 900} = \frac{850}{1500} = \frac{17}{30}$$

$$q = 1 - \frac{17}{30} = \frac{13}{30}$$

$$\text{S.E. } (p_1 - p_2) = \sqrt{\frac{17}{30} \times \frac{13}{30}\left(\frac{1}{600} + \frac{1}{900}\right)} = \sqrt{\frac{17}{30} \times \frac{13}{30} \times \frac{1500}{600 \times 900}} = 0.026$$

$$\frac{\text{Difference}}{\text{S.E.}} = \frac{0.667 - 0.44}{0.026} = 8.7$$

Since the difference is ore than 2.58 S.E. at 1% level of significance, the hypothesis is rejected. Hence there is a significant difference in the habit of smoking of the two cities.

3.7 SAMPLING OF VARIABLES

Q16. What do you understand by Sampling of Variable? State the objectives of sampling of variables.

(OR)

Define the term Sampling of Variable.

Ans :

Meaning

Sampling of variables such as height, weight, etc., which may take any value. It shall not, therefore, be possible for us to classify each member of a sample under one of two heads, success (or) failure. The values of the variable given by different trials will be spread over a range, which will be unlimited limited by practical considerations, as in the case of weight of people or which by theoretical considerations as in the case of correlation coefficient which cannot lie outside the range +1 to –1.

Objectives

(i) To compare observation with expectation and to see how far the deviation of one from the other can be attributed to fluctuations of sampling.

(ii) To estimate from samples some characteristic of the population, such as the mean of a variable and

(iii) To gauge the reliability of our estimates.

3.8 LARGE SAMPLES

Q17. Define the term large sample. What are the assumptions of large samples.

Ans : **(Imp.)**

Meaning

If the sample size, n > 30, then we consider such samples as large samples. The tests of significance used in large samples are different from those used in small samples because small samples fail to satisfy the assumptions under which large sample analysis is done. If n is large, the distributions, such as Binomial, Poisson, Chi-square etc. are closely approximated by normal distributions. Therefore, for large samples, the sampling distribution of a statistic is approximately a normal distribution.

Suppose we wish to test the hypothesis that the probability of success in such trial is p. Assuming it to be true, the mean μ and the standard deviation σ of the sampling distribution of number of successes are np and $\sqrt{npq}$ respectively.

If x be the observed number of successes in the sample and Z is the standard normal variate then $Z = \frac{x-\mu}{\sigma}$.

Thus we have the following test of significance

(i) If $|Z| < 1.96$, the difference between the observed and expected number of success is not significant.

(ii) If $|Z| > 1.96$, the difference is significant at 5% level of significance.

(iii) If $|Z| > 2.58$, the difference is significant at 1% level of significance.

Assumptions

The following are the assumptions under which significance tests are applie:

- The random sampling distribution of statistic has the properties of the normal curve. This may not hold good in case of small samples.
- Values (i.e., statistic) given by the samples are sufficiently close to the population values (i.e., parameters) and can be used in its place for calculating the standard, error (S. E.) of the estimate.

3.8.1 Difference between Large and Small Samples

Q18. What are the differences between Large and Small Samples?

(OR)

Compare and contrast Large and Small Samples.

Ans : **(Imp.)**

S.No.	Large sample	S.No.	Small sample
1.	The sample size is greater than 30.	1.	The sample size is 30 or less than 30
2.	The value of a statistic obtain from the sample can be taken as an estimate of the population parameter.	2.	The value of a statistic obtain from the sample can not be taken as an estimate of the population parameter.
3.	Normal distribution is used for testing.	3.	Sampling distribution like t, F etc. are used for testing.

3.8.2 Estimating Population Mean

Q19. Explain various formulae for estimating the appropriate size of sample.

Ans :

When the width of the confidence interval is specified, the size of the sample corresponding to this interval can be estimated as follows,

(i) Sample size for estimating a population mean

For infinite population,

$$n = \frac{z_c^2 \times \sigma^2}{E^2}$$

For Finite Population,

$$n = \frac{z_c^5 \times \sigma^2 \times N^2}{(N-1)E^2 + z^2\sigma^2}$$

Where,

E – Maximum allowable error,

$$E = Z_e \frac{\sigma}{\sqrt{n}}$$

z_c – z value for given confidence level

P – Population proportion of success

q – Population proportion of failure

n – Sample size

N – Finite population size.

z_c and E must be specified by the researcher.

(ii) Sample size for estimating a population proportion

For infinite population,

$$n = \frac{z_c^2 pq}{E^2}$$

For finite population,

$$n = \frac{z_c^2 pqN}{(N-1)E^2 + z_c^2 pq}$$

Where,

E – Maximum allowable error,

$$E = z_c\sqrt{\frac{pq}{n}}$$

z_c – z value for given confidence level

p – Population proportion of success

q – Population proportion of failure

σ – Population standard deviation

n – Sample size

N – Finite population size.

z_c and E must be specified by the researcher.

3.8.3 Testing

3.8.3.1 Significance of Mean

Q20. Explain the hypothesis concerning one mean for large samples.

Ans : **(Imp.)**

Let a random sample of size n(n > 30) has the sample mean $\bar{x}$, and population mean μ. Also the population mean μ has a specified value μ_0.

Working Rule

(i) The Null Hypothesis

$H_0 : \bar{x} = m$ i.e., "there is no significance difference between the sample mean and population mean" or the sample has been drawn from the parent population.

(ii) The Alternative Hypothesis : is

(i) $H_1 : \bar{x} \neq \mu\ (\mu \neq \mu_0)$ or

(ii) $H_1 : \bar{x} > \mu\ (\mu > \mu_0)$ or

(iii) $H_1 : \bar{x} \neq \mu\ (\mu < \mu_0)$

(iii) Level of Significance

Set the level of significance α.

(iv) The Test Statistic

We have the following two cases.

Case (i)

When the standard deviation σ of population is known.

In this case, standard Error of Mean, S.E. $(\bar{x}) = \frac{\sigma}{\sqrt{n}}$, where n = sample size, σ = s.d. of population.

$\therefore$ The test statistic is

$$z = \frac{\bar{x} - \mu}{S.E.(\bar{x})} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$, where μ is the sample mean.

Case (ii)

When the standard deviation σ of population is not known.

In this case, we take s, the s.d. of sample to compute the S.E. of mean

$$\therefore \quad S.E.\ (\bar{x}) = \frac{s}{\sqrt{n}}$$

Hence the test statistic is

$$z = \frac{x - \mu}{S.E.(\bar{x})} = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

(v) Find the critical value z_α of z at the level of significance α from the normal table.

(vi) Decision

(a) If $|z| < z_\alpha$, we accept the Null Hypothesis H_0.

(b) If $|z| > z_\alpha$, we reject the Null Hypothesis H_0.

The rejection rule for $H_0 : \bar{x} = \mu$ (or $\mu = \mu_0$) is given below :

Table : Critical values of z

Level of significance α	1%	5%	10%
Critical region for $\mu \neq \mu_0$	$\|z\| > 2.58$	$\|z\| > 1.96$	$\|z\| > 1.645$
Critical region for $\mu > \mu_0$	$z > 2.33$	$z > 1.645$	$z > 1.28$
Critical region for $\mu < \mu_0$	$z < -2.33$	$z < -1.645$	$z < -1.28$

Note

1. We reject Null Hypothesis H_0 when $|z| > 3$ without mentioning any level of significance.

The test statistic is, $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ where σ is the S.D of the population.

If the population S.D is not known, then use the statistic

$$z = \frac{\bar{x} - \mu}{S / \sqrt{n}}$$ where S is the sample S.D.

2. The values $\bar{x} \pm 1.96 \dfrac{\sigma}{\sqrt{n}}$ are called 95% fiducial limits or confidence limits for the mean of the population corresponding to the given sample.

Similarly, $\bar{x} \pm 2.58 \dfrac{\sigma}{\sqrt{n}}$ (or) [$\bar{x}$ – 2.58 (S.E. of $\bar{x}$), $\bar{x}$ + 2.58 (S.E of $\bar{x}$)] are called 99% confidence limits

and $\bar{x} \pm 2.33 \dfrac{\sigma}{\sqrt{n}}$ are called 98% confidence limits.

PROBLEMS

6. The mean lifetime of a sample of 100 light tubes produced by a company is found to be 1,560 hours with a population standard deviation of 90 hours. Test the hypothesis that the mean lifetime of the tubes produced by the company is 1580 hours.

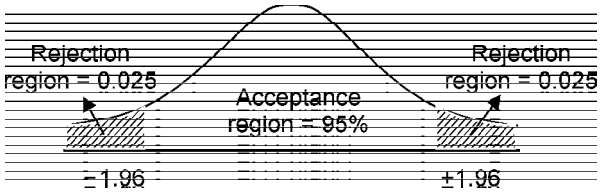
Sol :

The null hypothesis is that there is no significant difference between the sample mean and hypothetical population mean. The population mean has a specified value.

Null Hypothesis (H_0) : $m = m_0$

Alternative Hypothesis (H_1) : $m \neq m_0$

Let a = 0.05



Then the test statistics

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$Z = \frac{1560 - 1580}{90/\sqrt{100}} = -\frac{20}{90/10} = -\frac{20}{9} = -2.22$$

Since it is two-tailed test, the critical value of Z_a = ± 1.96 for a two-tailed test at α = 5% level of significance.

Conclusion : Since the computed value of Z = –2.22 in the rejection region, we reject the null hypothesis. Hence the mean lifetime of the tubes produced by the company may not be 1,580 hours.

7. The mean weight of 200 male students in a college is 62 kgs with a standard deviation of 4 kgs. Test the hypothesis that the mean weight in the population is greater than 58 kgs. Use a = 1%.

Sol :

The null hypothesis is that the population mean has a specified value, i.e.,

H_0 : m = 58 kgs (= m_0)

against H_1 : m > 58 kgs

a = 0.01

the test statistic

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$Z = \frac{60 - 58}{4/\sqrt{200}} = \frac{2}{4/14.14} = \frac{2}{0.28} = 7.14$$

The critical value of Z_a = 2.33 for a right - tailed at 1% level of significance.

Conclusion

Since computed value of Z = 7.14 falls in the rejection region, we reject the null hypothesis. Hence, the mean weight of the students is greater than 58 kgs.

8. A bakery is a city was fined Rs.1210 for selling cakes that were under weight. The city prosecutor collected 200 cakes from the bakery, to test the average weight of the cake. It is known that the population standard deviation (weight of the cake) is 2 kgs. The sample (n = 200) they obtained has a mean weight is 23.75 kgs. Will you conclude that the bakery was selling cakes that were under weight. The baker was expected to sell cake of weight 24 kgs.

Sol :

Null hypothesis

There is no significance difference between sample mean and hypothetical population mean (m = 24 kgs)

H_0 : m = 24 and

H_1 : m < 24 (left tailed test)

Let a = 0.01. The test statistic

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$Z = \frac{23.75 - 24}{2/\sqrt{200}} = \frac{-0.25}{0.141} = -1.77$$

Conclusion

Since the computed value of $Z = -1.7678$ falls in the acceptance region, we may accept the null hypothesis. Hence the bakery was selling cakes of correct weight of 24 kgs, only.

3.8.3.2 Significance of the Difference between Means of Two Samples

Q21. Explain briefly about hypothesis concerning two mean large samples.

Ans : **(Imp.)**

Let $\bar{x}_1$ and $\bar{x}_2$ be the sample means of two independent large random samples sizes n_1 and n_2 drawn from two populations having means μ_1 and μ_2 and standard deviations σ_1 and σ_2. To test whether the two population means are equal.

Null Hypothesis is H_0 : $\mu_1 = \mu_2$

Alternative Hypothesis is H_1 : $\mu_1 \neq \mu_2$

S.E. of $(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, where σ_1 and σ_2 are the S.D. of the two populations

To test whether there is any significant difference between $\bar{x}_1$ and $\bar{x}_2$, we have to use statistic

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\text{S.E. of } (\bar{x}_1 - \bar{x}_2)} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

where $\delta = \mu_1 - \mu_2$ (= given constant)

If $\delta = 0$, the two populations have the same means.

If $\delta \neq 0$ the two populations are different.

Under $H_0 : \mu_1 = \mu_2$, the test statistic becomes $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

is approximately normally distributed with mean 0 and S. D. 1.

Note

If the samples have been drawn from the population with common S. D. σ, then $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Hence $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma \cdot \sqrt{\dfrac{1}{n_1^2} + \dfrac{1}{n_2^2}}}$

is normally distributed with mean zero and standard deviation one.

If σ is not known we can use an estimate of σ^2 given by $\sigma^2 = \dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$

Rejection Rule for H_0 : $\mu_1 - \mu_2$

(i) If $|z| > 1.96$, reject H_0 at 5% level of significance.

(ii) If $|z| > 2.58$, reject H_0 at 1% level of significance.

(iii) If $|z| > 1.645$, reject H_0 at 10% level of significance.

(iv) If $z > 3$ then either the samples have not been drawn from the some population or the sampling is not simple.

Otherwise accept H_0.

Note :

If the two samples are drawn from two population with unknown standard deviations σ_1^2 and σ_2^2, then σ_1^2 σ_2^2 can be replaced by sample variances S_1^2 and S_2^2 provided both the samples n_1 and n_2 are large.

In this case, the test statistic is $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$

PROBLEMS

9. A factory uses process A and process B for manufacturing the same item. The average weight in a sample of 250 items produced from process A is found to be 120 grams with a standard deviation of 12 grams, while for the process B, average weight in a sample of 400 items is found to be 124 grams with a standard deviation of 14 grams. Test whether the average weights of both the processes are same?

Sol : **(Imp.)**

It is a test on large sample of two mean

A	B
$n_1 = 250$	$n_2 = 400$
$\bar{x}_1 = 120$	$\bar{x}_2 = 124$
$\sigma_1 = 12$	$\sigma_2 = 14$
$\sigma_1^2 = 144$	$\sigma_2^2 = 196$

(i) Null Hypothesis

$H_0 = M_1 = M_2$

There is no significant difference in average weight for manufacturing item in A and B.

(ii) Alternative Hypothesis

$H_0 : M_1 \neq M_2$ (two tailed 'Z' test)

The average weight for manufacturing item in A and B are not same.

(iii) Computing Test Statistic

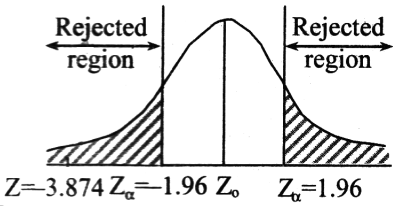
$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{120 - 124}{\sqrt{\frac{144}{250} + \frac{196}{400}}}$$

$$Z = \frac{-4}{1.03247}$$

$$Z = -3.87420$$

(iv) Level of significance at 5% i.e., $\alpha = 0.05$

(v) Critical Value $Z_\alpha = 196$



Decision

Null Hypothesis lies in rejection region.

$\therefore$ Accept H_a at 5% level of significance

i.e., the average weight for manufacturing items in A and B are equal or same i.e., $M_1 \neq M_2$

10. You are given the position in a factory before and after settlement of an industrial dispute. Comment on the gains or losses from the point of view of workers and that of management.

	Before	After
No. of workers	4800	4700
Mean wages (Rs.)	900	950
Median wages (Rs.)	960	900
Standard deviation (Rs.)	240	200

Sol :

Let the null hypothesis be that there is no change in wages before and after settlement.

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_{s_1}^2}{n_1} + \frac{\sigma_{s_2}^2}{n_2}}}$$

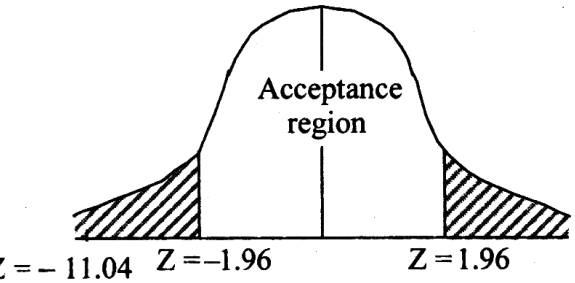
Given $\bar{x}_1 = 900$; $\sigma_{s_1} = 240$; $n_1 = 4{,}800$

$\bar{x}_2 = 950$; $\sigma_{s_2} = 200$; $n_2 = 4{,}700$

$$\therefore \quad Z = \frac{900 - 950}{\sqrt{\frac{240^2}{4800} + \frac{200^2}{4700}}}$$

$$= \frac{-50}{\sqrt{12 + 8.51}} = \frac{-50}{4.53} = -11.04$$

Rejection region for $\alpha = 0.05$ (assumed) is $Z > 1.96$ and $Z < -1.96$.



Calculated value of Z = – 11.04

Since calculated value of Z falls in the rejection region, H_0 is rejected and H_1 is accepted i.e., is a difference in wages in wages before and after the settlement.

Hence, we can conclude that, from the point of view of workers, there are gains from the point of view of management, there are losses.

3.8.3.3 Significance of the difference between the Standard Deviations of Two Samples

Q22. Explain test of significance for difference of standard deviations with example.

Ans : **(Imp.)**

Let S_1 and S_2 be the standard deviations of two independent random samples of sizes n_1 and n_2 for two populations with standard deviations σ_1 and σ_2 respectively.

Null Hypothesis, $H_0 : \sigma_1 = \sigma_2$ i.e., the sample standard deviations don't differ significantly.

Test statistic, $Z = \frac{s_1 - s_2}{S.E(s_1 - s_2)}$ –N(0, 1) (for large samples)

In sampling from a large population.

Sampling distribution of s is,

$$\text{Var}(S) = \frac{\sigma^2}{2n}$$

or $$\text{S.E}(S) = \frac{\sigma^2}{\sqrt{2n}}$$

$$\therefore \quad \text{Var}(s_1 - s_2) = \text{Var}(S_1) + \text{Var}(S_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma^2}{2n_2}}$$

As samples are independent, covariance term is discarded.

$$\therefore \quad Z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}$$

σ_1^2 and σ_2^2 are usually unknown and for large samples, we use estimates given by their corresponding samples variance.

Hence, the test statistic reduces to,

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \sim N(0, 1) \text{ (for large samples)}$$

PROBLEMS

11. There are 5,000 students in a college. Out of 1,00,000 in the whole university, in a study 1,000 were found smokers in the college and 5,000 in the whole university. Is there a significant difference between the proportion of smokers in the college and university ?

Sol :

H_0 : Null Hypothesis

There is no significant difference between the proportion of smokers in the college and whole university.

H_1: Alternative Hypothesis

There is a significant difference between the proportion of smokers in the college and whole university.

Proportion of smokers in the college is,

$$\Rightarrow \quad \frac{1,000}{5,000} = 0.2$$

Proportion of smokers in the whole university is $\frac{5,000}{1,00,000} = 0.05$

Difference between the two proportions is,

$\Rightarrow$ 0.2 – 0.05 = 0.15

Proportion of smokers in the whole university is represented by $P_0 = 0.05$.

$q_0 = 1 - P_0$
$= 1 - 0.05$
$= 0.95$
$n_1 = 5{,}000,\ n_1 + n_2 = 1{,}00{,}000$
$n_2 = 1{,}00{,}000 - 5{,}000$
$= 95{,}000$

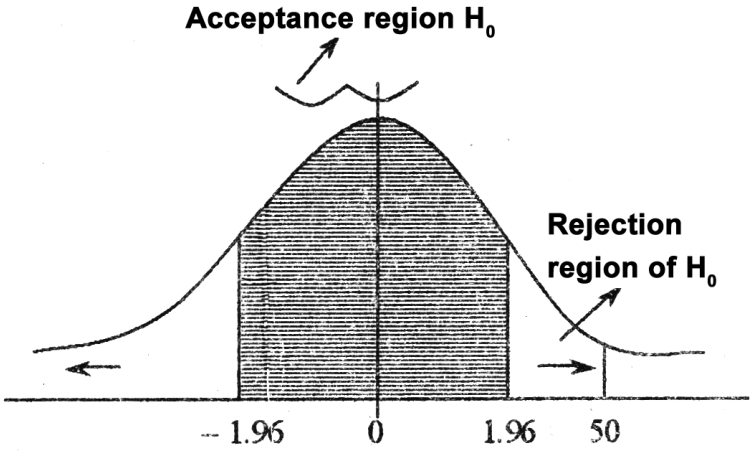
$\therefore$ S.E of difference $= \sqrt{P_0 q_0 \times \frac{n_2}{n_1(n_1 + n_2)}}$

$$= \sqrt{0.05 \times 0.95\left[\frac{95{,}000}{5{,}000(5{,}000 + 95{,}000)}\right]} = \sqrt{0.0475\left(\frac{95{,}000}{5{,}000(1{,}00{,}000)}\right)}$$

$$= \sqrt{0.0475\left(\frac{95{,}000}{50{,}00{,}000}\right)} = \sqrt{0.0475 \times 0.00019}$$

$$= \sqrt{0.000009} = 0.003$$

$$\frac{\text{Difference}}{\text{S.E}} = \frac{0.15}{0.003} = 50$$



Decision

5% level of significance is 'z' value 1.96. As 50 is more than 1.96, there is a significant difference between the proportion of smokers in the college and university.

3.9 SMALL SAMPLES

Q23. Explain briefly about small sample test ?

Ans :

When the size of sample is small (less than 30), the above tests are not applicable because the assumptions on which they are based generally do not hold good in case of small samples. In particular, it will no longer be possible for us to assume

(a) That the random sampling distribution of a statistic is approximately normal, and

(b) That values given by the sample data are sufficiently close to the population values and can used in their place for the calculation of the standard error of the estimate.

The removal of these a assumptions makes it necessary to use entirely new techniques to deal with the problems of small samples. The division between the theories of large and small samples is, therefore, a very real one, though it is not always easy to draw a precise line of demarcation. It should be noted that as a rule, the methods and the theory of small samples are applicable to large samples, though the reverse is not true.

3.9.1 't' test

Q24. Define 't' distribution. Explain the properties and applications of 't' distribution.

Ans : **(Imp.)**

Meaning

When population standard deviation (σ_p) 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,

$$t = \frac{(\bar{x} - \mu)}{\frac{\sigma_s}{\sqrt{n}}}$$

Where, $\sigma_s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$

i.e., the sampling standard deviation.

Properties

The properties of student's t-distribution are as follows,

(i) The probability curve of t is symmetric, like in standard normal distribution (Z).

(ii) The t-distribution ranges from $-\alpha$ to α just as does a normal distribution.

(iii) The t-distribution is bell shaped and symmetrical around mean zero, like normal distribution.

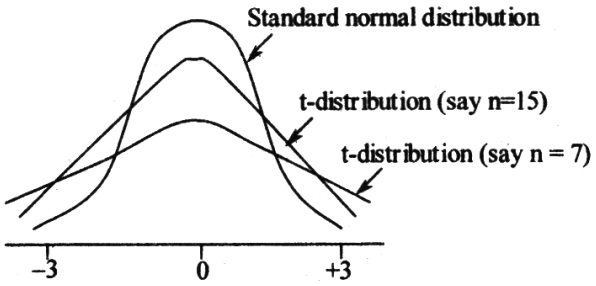
(iv) The shapes of the t-distribution changes as the sample size changes (the number of degrees of freedom changes) whereas it is same for all sample sizes in z-distribution.

(v) The variance of t-distribution is always greater than one and is defined only when $n \geq 3$.

(vi) The t-distribution is more of platykurtic (less packed at centre and higher in tails) than the normal distribution.

(vii) The t-distribution has a greater dispersion than the normal distribution. As n becomes larger, the t-distribution approaches the standard normal distribution.

(viii) There is a family of t-distribution one for each sample size whereas, there is only one standard normal distribution.



Applications

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.

3.9.2 Significance of the Mean

3.9.2.1 Fixing Fiducial Limits to Population Mean

Q25. Explain the test concerning the significance of single mean of a random sample. What are the Fixing Fiducial Limits to Population Mean?

Ans : **(Imp.)**

In determining whether the mean of a sample drawn from a normal population deviates significantly from a stated value (the hypothetical value of the populations mean), when variance of the population is unknown we calculate the statistic :

$$t = \frac{(\bar{X} - \mu)\sqrt{n}}{S}$$

where $\bar{X}$ = the mean of the sample

μ = the actual (or) hypothetical mean of the population

n = the sample size

S = the standard deviation of the sample

$$S = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}} \text{ or } S = \sqrt{\frac{\sum d^2 - n(\bar{d})^2}{n-1}} = \sqrt{\frac{1}{(n-1)}\left[\sum d^2 - \frac{(\sum d)^2}{n}\right]}$$

where d = deviation from the assumed mean.

If the calculated value of |t| exceeds $t_{0.05,}$ we say that the difference between $\bar{X}$ and μ is significant at 5% level, if it exceeds $t_{0.01}$, the difference is said to be significant at 1% level. If $|t| < t_{0.05}$, we conclude that difference between $\bar{X}$ and μ is not significant and hence the sample might have been drawn from a population with mean = μ.

Fiducial Limits of Population Mean

Assuming that the sample is a random sample from a normal population of unknown mean the 95% fiducial limits of the population mean (μ) are :

$$\bar{X} \pm \frac{S}{\sqrt{n}} t_{0.05}$$

and 99% limits are $$\bar{X} \pm \frac{S}{\sqrt{n}} t_{0.05}$$

PROBLEMS

12. The mean price of shares of Andhra Bank during 2004 was `64. In the year 2005 the mean price of Andhra Bank for 20 randomly selected days is found to be `84. with a standard deviation of `4.50. Test whether there is a significant difference in the price of shares for the two years of 5% significance level.

Sol :

Population standard deviation not known. Sample size small (less than 30) t-test used.

Test for a specified mean,

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

Let the null hypothesis be that there is no significant difference in the prices of shares for two years.

Given,

$m = 64, n = 20$ days

$\bar{X} = 84, S = 4.5$

Computed test statistic is,

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}} = \frac{84 - 64}{4.5/\sqrt{20}}$$

$$= \frac{20}{1.0062} = 19.877$$

Degrees of freedom, d.f = n – 1 = 20 – 1 = 19

For d.f = 19, $t_{0.05}$ = 2.093 (for two tailed test)

Decision

As calculated value of t(i.e., 19.877) is more than table value i.e., 2,093, null hypothesis is rejected. There is difference in price of shares for 2 years.

13. A random sample of size 16 has 53 as mean. The sum of the squares of the deviations taken from mean is 135. Can this sample be reagarded as taken from the population having 56 as mean ? Obtain 95% and 99% conference limits of the mean of the population. (for v = 15, $t_{0.05}$ = 2.13 for v = 15 t_0.01 = 2.95).

Sol :

Let us take the hypothesis that there is no significant difference between the sample mean and hypothetical population mean. Applying t test :

$$t = \frac{\bar{X} - \mu}{S}\sqrt{n}$$

$\bar{X} = 53, \mu = 56, n = 16, \sum(X - \bar{X})^2 = 135$

$$S = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} = \sqrt{\frac{135}{15}} = \sqrt{9} = 3$$

$$t = \frac{|53-56|}{3}\sqrt{16} = \frac{3\times 4}{3} = 4$$

$v = 16 - 1 = 15$. For $v = 16$, $t_{0.05} = 2.13$

The calculated value of t is more than the table value. The hypothesis is rejected. Hence, the sample has not come from a population having 56 as mean.

95% confidence limits of the population mean

$$\bar{X} \pm \frac{S}{\sqrt{n}}\ t_{0.05}$$

$$= 53 \pm \frac{3}{\sqrt{16}} \times 2.13$$

$$= 53 \pm 1.6 = 51.4 \text{ to } 54.6$$

99% confidence limit of the population mean

$$\bar{X} \pm \frac{S}{\sqrt{n}}\ t_{0.05}$$

$$= 53 \pm \frac{3}{\sqrt{16}} \times 2.95$$

$$= 53 \pm \frac{3}{4} \times 2.95$$

$$= 53 \pm 2.212 = 50.788 \text{ to } 55.212$$

14. The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data

Item :	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Life in '000 hour :	**4.2**	**4.6**	**3.9**	**4.1**	**5.2**	**3.8**	**3.9**	**4.3**	**4.4**	**5.6**

Can we accept the hypothesis that the average life time of bullbs is 4,000 hours.

Sol : **(Imp.)**

Let us take the hypothesis that there is no significant difference in the sample mean and the hypothetical population mean. Applying the t-test.

$$t = \frac{\bar{X}-\mu}{S}\sqrt{n}$$

Calculation of $\bar{X}$ and S

X	$(X - \bar{X})$	$(X - \bar{X})^2$
4.2	– 0.2	0.04
4.6	+ 0.2	0.04
3.9	– 0.5	0.25
4.1	– 0.3	0.09
5.2	+ 0.8	0.64
3.8	– 0.6	0.36
3.9	– 0.5	0.25
4.3	– 0.1	0.01
4.4	0	0
5.6	+ 1.2	1.44
$\Sigma X = 44$		$\Sigma(X - \bar{X})^2 = 3.12$

$$S = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}} = \sqrt{\frac{3.12}{9}} = 0.589$$

$$t = \frac{4.4 - 4}{0.589}\sqrt{10} = \frac{0.4 \times 3.162}{0.589} = 2.148$$

The calculated value of t is less than the table value. The hypothesis is accepted. The average life time of the bulbs could be 4000 hours.

5.9.3 Significance of the Difference between Two Independent Means

Q26. Explain the test concerning the significance of two independent means of a random sample.

Ans :

Given two independent random samples of size n_1 and n_2 with means $\bar{X}_1$ and $\bar{X}_2$ and standard deviations S_1 and S_2 we may be interested in testing the hypothesis that the samples come from the same normal population. To carry out the test, we calculate the statistic as follows :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

where $\bar{X}_1$ = mean of the sample

$\bar{X}_2$ = mean of the second sample

n_1 = number of observations in the first sample

n_2 = number of observations in the second sample

S = combined standard deviation

The value of S is calculated by the following formula :

$$S = \sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

When the actual means are in fraction the deviations should be taken from assumed means. In such a case combined standard deviation is obtained by applying the following formula :

$$S = \sqrt{\frac{\sum(X_1 - A_1)^2 + \sum(X_2 - A_2)^2 - n_1(\bar{X} - A_1)^2 - n_2(\bar{X}_2 - A_2)^2}{n_1 + n_2 - 2}}$$

A_1 = Assumed mean of the first sample

A_2 = Assumed mean of the second sample

$\bar{X}_1$ = Actual mean of the first sample

$\bar{X}_2$ = Actual mean of the second sample

The degrees of freedom = $(n_1 + n_2 - 2)$

When we are given the number of observations and standard deviation of the two samples the pooled estimate of standard deviation can be obtained as follows :

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

If the calculated value of t be $> t_{0.05}$ ($t_{0.01}$), the difference between the sample means is said to be significant at 5% (1%) level of significance otherwise the data are said to be consistent with the hypothesis.

PROBLEMS

15. Two types of drugs were used on 5 and 7 patients for reducing their weight.

Drug A was imported and drug B indigenous. The decrease in the weight after using the drugs for six months was as follows :

Drug A	**10**	**12**	**13**	**11**	**14**		
Drug B	**8**	**9**	**12**	**14**	**15**	**10**	**9**

Is there a significant difference in the efficacy of the two drugs ? if not, which drug should you buy ?)For v = 10, $t_{0.05}$ = 2.223).

Sol :

Let us take the hypothesis that there is no significant difference in the efficacy of the two drugs. Applying t-test :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

X_1	$(X_1 - \bar{X}_1)$	$(X_1 - \bar{X}_1)^2$	X_2	$(X_2 - \bar{X}_2)$	$(X_2 - \bar{X}_2)^2$
10	– 2	4	8	– 3	9
12	0	0	9	– 2	4
13	+ 1	1	12	+ 1	1
11	– 1	1	14	+ 3	9
14	+ 2	4	15	+ 4	16
			10	– 1	1
			9	– 2	4
$\Sigma X_1 = 60$	$\Sigma(X_1 - \bar{X}_1)$ = 10	$\Sigma(X_1 - \bar{X}_1)^2$ = 10	$\Sigma X_2 = 77$	$\Sigma(X_2 - \bar{X}_2)$ = 10	$\Sigma(X_2 - \bar{X}_2)^2$ = 44

However, it is advisable to take account of bias

$$\bar{X}_1 = \frac{\sum X_1}{n_1} = \frac{60}{5} = 12;\ \bar{X}_2 = \frac{\sum X_2}{n_2} = \frac{77}{7} = 11$$

$$S = \sqrt{\frac{\sum (X_1 - X_1)^2 + \sum (X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{10 + 44}{5 + 7 - 2}} = \sqrt{\frac{54}{10}} = 2.324$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$\frac{12 - 11}{2.324}\sqrt{\frac{5 \times 7}{5 + 7}} = \frac{1}{2.324} \times 1.708 = \frac{1.708}{2.324} = 0.735$$

$$v = n_1 + n_2 - 2 = 5 + 7 - 2 = 10$$

For $v = 10,\ t_{0.05} = 2.228$

The calculated value of t is less than the table value, the hypothesis accepted. Hence, there is no significance in the efficacy of two drugs. Since drug B is indigenous and there is no difference in the efficacy of imported and indigenous drug, we should buy indigenous drug. i.e., B.

3.9.4 Significance of the Difference between Two Dependent Means (Paired 't' Test)

Q27. Explain the test concerning the significance of two dependent means of a random sample.

(OR)

Discuss in detail about Paired Sample 't' Test

Ans :

Two samples are said to be dependent when the elements in one sample are related to those in the other in any significant or meaningful manner. In fact, the two samples may consist of pairs of observations made on the same object, individual or, more generally, on the same selected population elements. When samples are dependent they comprise the same number of elementary units.

$$t = \frac{\bar{d}\sqrt{n}}{S}$$

where $\bar{d}$ = the mean of the differences

S = the standard deviation of the differences

The value of S is calculated as follows :

$$S = \sqrt{\frac{\sum(d-\bar{d})^2}{n-1}} \text{ or } \sqrt{\frac{\sum d^2 - n(\bar{d})^2}{n-1}}$$

It should be noted that t is based on n – 1 degrees of freedom.

PROBLEMS

16. To verify whether a course in accounting improved performance, a similar test was given to 12 participants both before and after the course. The original marks recorded in alphabetical order of the participants were 44, 40, 61, 52, 32, 44, 70, 41, 67, 72, 53 and 72. After the course, the marks were in the same order, 53, 38, 69, 57, 46, 39, 73, 48, 73, 74, 60 and 78. Was the course useful ?

Sol :

Let us take the hypothesis that there is no difference in the marks obtained before and after the course, i.e., the course has not been useful.

Applying t-test (difference formula) :

$$t = \frac{\bar{d}\sqrt{n}}{S}$$

Participants	Before (1st Test)	After (2nd Test)	(2nd-1st test) d	d^2
A	44	53	+ 9	81
B	40	38	– 2	4
C	61	69	+ 8	64
D	52	57	+ 5	25
E	32	46	+ 14	196
F	44	39	– 5	25
G	70	73	+ 3	9
H	41	48	+ 7	49
I	67	73	+ 6	36
J	72	74	+ 2	4
K	53	60	+ 7	49
L	72	78	+ 6	36
			$\Sigma d = 60$	$\Sigma d^2 = 578$

$$\bar{d} = \frac{\sum d}{n} = \frac{60}{12} = 5$$

$$S = \sqrt{\frac{\sum d^2 - n(\bar{d})^2}{n-1}}$$

$$= \sqrt{\frac{578 - 12(5)^2}{12-1}} = \sqrt{\frac{278}{11}}$$

$$= 5.03$$

$$t = \frac{5 \times \sqrt{12}}{5.03} = \frac{5 \times 3.464}{5.03} = 3.443$$

$v = n - 1 = 12 - 1 = 11$; For $v = 11$, $t_{0.05} = 2.201$

The calculated value of t is greater than the table value. The hypothesis is rejected. Hence the course has been useful.

17. Eleven sales executive trainees are assigned selling jobs right after their recruitment. After a fortnight they are withdrawn from their field duties and given a month's training for executives sales. Sales executed by them in thousands of rupees before and after the training in the same period are listed below.

Sales (000 `) (Before training)	**23**	**20**	**19**	**21**	**18**	**20**	**18**	**17**	**23**	**16**	**19**
Sales (000 `) (After training)	**24**	**19**	**21**	**18**	**20**	**22**	**20**	**20**	**23**	**20**	**27**

Do these data indicate that the training has contributed to their performance?

Sol : **(Imp.)**

Let the null hypothesis be that the training program not contributed to sales.

$H_0 : \mu_1 = \mu_2$

Let the alternate hypothesis be training program has contributed to sales.

$H_a : \mu_1 < \mu_2$

Sample size small population variance not known.

Paired t-test used,

$$t = \frac{\bar{d}}{\sqrt{s^2 / n}} \text{ (or) } \frac{\bar{d}}{S / \sqrt{n}}$$

$$S = \frac{1}{n-1}\left[\Sigma d^2 \frac{(\Sigma d)^2}{n}\right] \text{ (or)}$$

$$S = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n-1}}$$

Sales (in thousands) `:Difference

Sl.No.	Before Training (1)	After Training (2)	d = (1) – (2)	d^2
1	23	24	–1	1
2	20	19	1	1
3	19	21	–2	4
4	21	18	3	9
5	18	20	–2	4
6	20	22	–2	4
7	18	28	–2	4
8	17	20	–3	9
9	23	23	0	0
10	16	20	–4	16
11	19	27	–8	64
			$\Sigma d = -20$	$\Sigma d^2 = 116$

Mean, $\bar{d} = \frac{\Sigma d}{n} = \frac{-20}{11} = -1.82$

$$S^2 = \frac{1}{n-1} = \left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]$$

$$= \frac{1}{11-1} = \left[116 - \frac{(20)^2}{11}\right] = \frac{1}{10}\left(116 - \frac{400}{11}\right)$$

$$= \frac{1}{10} \times \frac{876}{11} = \frac{876}{110}$$

$S^2 = 7.96$

$$t = \frac{\bar{d}}{\sqrt{S^2/n}} = \frac{-1.82}{\sqrt{78.96/11}}$$

$$= \frac{-1.82}{\sqrt{0.72}} = \frac{-1.82}{0.85}$$

$= -2.14$

Critical Value

Degrees of freedom, (d.f) = n – 1
= 11 – 1
= 10

[**Note:** Assuming level of significance as 5%].

Decision

For degrees of freedom (d.f) = 10 the table value of (5%) $t_{0.05}$ for two - tailed test 2.228. The table value is more than computed value hence null hypothesis is accepted that is training program has not contributed to sales.

Short Questions and Answers

1. Hypothesis.

Ans :

Definitions

Hypothesis is defined in different ways by different people, some of them are as follows,

i) **According to James E. Greighton,** "It is a tentative supposition or provisional guess which seems to explain the situation under observation".

ii) **According to Bruce W. Titckman,** "A hypothesis then could be defined as an expectation about events based on generalization of the assumed relationship between variables".

2. Features of a hypothesis

Ans :

i) Hypothesis is conceptual in nature as few conceptual elements are included in its framework.

ii) It is basically a verbal statement in a declarative type, which has a verbal expression of ideas and concepts which are ready for empirical inspection.

iii) It holds empirical referent which highlights the tentative relationship between two (or) more variables.

iv) It is future oriented in nature and is associated with the future verification and does not take into consideration the historical facts and informations.

3. Importance of hypothesis

Ans :

i) Hypothesis basically acts as the eyes of investigator by supporting him in finding out answers for tentatively following generalization.

ii) Hypothesis usually emphasizes upon research. In absence of research hypothesis there would be a more waiting.

iii) It acts as an important link between theory and the investigation.

4. Hypothesis

Ans :

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.

5. Directional Hypothesis

Ans :

Directional hypothesis, basically emphasizes upon drawing an expected direction for the existing relationship between different variables.

6. Sampling of Attributes

Ans :

The sampling of attributes may, therefore, be regarded as the drawing of samples from a population whose members possess the attribute A or not A.

For example, in the study of attribute 'Literacy' a sample may be taken and people classified as literates and illiterates. With such data the binomial type of problem may be formed. The selection of an individual on sampling may be called 'event', the appearance of attribute A may be taken as 'success' and its non-appearance as 'failure'. Thus, if out of 1,000 people selected for the samples, 100 are found literates, and 900 illiterates, we would say that the sample consists of 1000 units out of which 100 are successes and 900 failures. The probability of success or p = 100 / 1,000 or 0.1 and the probability of failure or q = 900 / 1,000 = 0.9 so that p + q = 0.1 + 0.9 = 1.

7. Sampling of Variable.

Ans :

Meaning

Sampling of variables such as height, weight, etc., which may take any value. It shall not, therefore, be possible for us to classify each member of a sample under one of two heads, success (or) failure. The values of the variable given by different trials will be spread over a range, which will be unlimited limited by practical considerations, as in the case of weight of people or which by theoretical considerations as in the case of correlation coefficient which cannot lie outside the range +1 to –1.

8. Large sample.

Ans :

Meaning

If the sample size, n > 30, then we consider such samples as large samples. The tests of significance used in large samples are different from those used in small samples because small samples fail to satisfy the assumptions under which large sample analysis is done. If n is large, the distributions, such as Binomial, Poisson, Chi-square etc. are closely approximated by normal distributions. Therefore, for large samples, the sampling distribution of a statistic is approximately a normal distribution.

9. Compare and contrast Large and Small Samples.

Ans : **(Imp.)**

S.No.	Large sample	S.No.	Small sample
1.	The sample size is greater than 30.	1.	The sample size is 30 or less than 30
2.	The value of a statistic obtain from the sample can be taken as an estimate of the population parameter.	2.	The value of a statistic obtain from the sample can not be taken as an estimate of the population parameter.
3.	Normal distribution is used for testing.	3.	Sampling distribution like t, F etc. are used for testing.

10. Small sample test.

Ans :

When the size of sample is small (less than 30), the above tests are not applicable because the assumptions on which they are based generally do not hold good in case of small samples. In particular, it will no longer be possible for us to assume

(a) That the random sampling distribution of a statistic is approximately normal, and

(b) That values given by the sample data are sufficiently close to the population values and can used in their place for the calculation of the standard error of the estimate.

Exercises Problems

1. In a random sample of 400 persons from a large population, 120 are females. Can it be said that males and females are in the ratio 5 : 3 in the population ? use 1% level of significance.

[Ans : Z = – 3.125; The males and females in the population are not in the ratio 5 : 3.]

2. A coin is tossed 900 times and heads appear 490 times. Does this res. support the hypothesis that the coin is unbiased.

[Ans : $H_0 : P = \frac{1}{2}$ (coin is unbiased), Z = 2.39, rejected at 5% level.]

3. A wholesaler in apples claims that only 4% of the apples supplied by him cr: defective. A random sample of 600 apples contained 36 defective apples. Te- the claim of the wholesaler.

[Ans : $H_0 : P = 0.04$; $H_1 : P > 0.04$, Z = 10, highly significant.]

4. In a sample of 500 people in Tamil Nadu 280 are tea drinkers and the rest ar: coffee drinkers. Can we assume that both coffee and tea are equally popular - this state at 1% level of significance.

[Ans : $H_0 : P = \frac{1}{2}$, $H_1 : P \neq \frac{1}{2}$, Z = 2.68, significant at 1% level.]

5. A manufacturer claimed that atleast 98% of the steel pipes which he supplic; to a factory conformed to specifications. An examination of a sample : 500 pieces of pipes revealed that 30 were defective. Test his claim a; . significance level 5%.

[Ans : $H_0 : P = 0.98$, $H_1 : P < 0.98$ (left-tail test), Z = 6.38, H_0 rejected.]

6. A sample of 100 workers in a large plant gave a mean assembly time of 294 seconds with a S.D. of 12 seconds in a time and motion study. Find a 95% confidence interval for the mean assembly time for all the workers in the plant.

[Ans : 291.64, 296.35.]

Choose the Correct Answer

1. The totality of the observation is called [a]
   (a) Population (b) Sample
   (c) Parameter (d) None
2. The statistical constants of the population are celled [b]
   (a) Statistic (b) Parameter
   (c) Sample statistic (d) None
3. The probability distribution of a statistic is called [b]
   (a) Normal distribution (b) Sampling distribution
   (c) Binomial distribution (d) None
4. The number of possible samples of size n out of N population units without replacement is [a]
   (a) N_{C_n} (b) N^n
   (c) $\frac{1}{N_{C_n}}$ (d) None
5. The probability of any one sample of size n being drawn out of N units is [a]
   (a) $\frac{1}{N_{C_n}}$ (b) N_{C_n}
   (c) $\frac{1}{N_{C_n}}$ (d) None
6. A population consisting of all real numbers is an example of [a]
   (a) An infinite population (b) An infinite population
   (c) Sample (d) None
7. The standard error of the statistic sample mean $(\bar{x})$ is [a]
   (a) $\frac{\sigma}{\sqrt{n}}$ (b) $\frac{\sigma^2}{\sqrt{n}}$
   (c) $\sqrt{\frac{\sigma}{n}}$ (d) None
8. _____ are the steps of hypothesis testing. [c]
   (a) Making statements and testing
   (b) Considering level of significance and critical yalue
   (c) Calculation of probability and comparing the probability
   (d) Application of one and two tailed tests

9. Z value for confidence level 95% is _____ . [a]

   (a) 1.96 (b) 1.645

   (c) 2.58 (d) 2.33

10. The _____ one of the important applications of t-distribution [b]

   (a) Only test of hypothesis

   (b) Test of hypothesis about the population mean

   (c) Test of hypothesis for large samples

   (d) Test of sample size

Fill in the blanks

1. _____ (or universe) is the aggregate or totality of statistical data forming a subject of investigation.
2. The procedure or rule to determine an unknown population parameter is called an _____.
3. Type-II error is denoted by beta known as _____.
4. _____ Error means rejection of hypothesis which should have been accepted
5. _____ theory is the study of relationships between a population and samples drawn from the population.
6. The _____ is the sampling distribution of the sampling means approaches a normal distribution as the sample size gets larger, no matter what the shape of the data distribution.
7. The standard deviation of sampling distribution of a statistic is known as its _____.
8. A point estimate is a single number which is used as an estimate of the unknown _____ parameter.
9. An _____ estimate of a population parameter is a statement of two values between which it is estimated that the parameter lies.
10. An estimate is a statement made to find an unknown _____ parameter.

ANSWERS

1. Population
2. Estimator
3. beta error
4. Type-I
5. Sampling
6. Central Limit Theorem
7. Standard error
8. Population
9. Interval
10. Population

One Mark Answers

1. Estimate

Ans :

An estimate is a statement made to find an unknown population parameter.

2. Efficiency

Ans :

The concept of efficiency refers to the sampling variability of an estimator. If two competing estimators are both unbiased, the one with the smaller variance (for a given sample size) is said to be relatively more efficient.

3. Non-Directional Hypothesis

Ans :

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.

4. Define 't' distribution.

Ans :

When population standard deviation (σ_p) 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.

5. Applications of t-distribution

Ans :

(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.

UNIT IV

Analysis of Variance, Association of Attributes & Chi-square Test:

Analysis of Variance: F-test: Meaning and Applications - ANOVA: Assumptions - Procedure - One way and two-way analysis of variance (including Problems).

Association of Attributes & Chi-Square Test: Association of Attributes: Meaning- Distinction between Correlation and Association - Methods of studying Association - Interpretation of results.

Chi Square Test: Definition - Conditions for applying Chi square test, Yates s correction– Uses and limitations of Chi square test – Chi square test for testing the independence of Attribute – Chi square test for goodness of fit (including problems).

4.1 F-TEST

4.1.1 Meaning and Applications

Q1. Define F-test. Explain properties and applications of F-test.

Ans :

Meaning

F-Test or F-distribution is a continuous probability distribution used when two different normal population, are sampled. Consider S_1^2 and S_2^2 as the sample variances of different random sample of sizes n_1 and n_2 respectively. These samples are drawn from two different normal population $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, where (μ_1, σ_1^2) and (μ_2, σ_2^2) denotes the mean and variances of Sf and S\ respectively.

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

Inorder to determine whether the samples (S_1^2, S_2^2) are drawn from two different population, having equal variances. It is necessary to compute the ratio of variances related to two independent random sample. This ratio is computed as,

If it is assumed that normal population have equal variance, then,

$$F = \frac{S_1^2}{S_2^2} \text{ if } S_1^2 > S_2^2$$

$$= \frac{S_2^2}{S_1^2} \text{ if } S_2^2 > S_1^2$$

Therefore, the sampling distribution of F can be written in the following form,

$$f(F) = \frac{K.F^{(V_1-2)/2}}{(V_1F + V_2)^{(V_1+V_2)/2}}$$

Where,

$$K = \int_0^\infty f(F)\, dF = 1,$$

Degree of freedom of n_1, $V_1 = n_1 - 1$

Degree of freedom of n_2, $V_2 = n_2 - 1$

The value of F is approximately equal to 1 when two sample variances are almost equal.

Properties

Following are the properties of F-distribution,

1. The F-distribution curve lies in only first quadrant (Q_1) and is unimodal.
2. The F-distribution is independent (free) of population parameter and depends only on the degree of freedom (i.e., V_1 and V_2) according to its order.
3. The F-distribution mode is less than unity (i.e., mode < 1).
4. In the F-distribution figure, $F_{1-\alpha}$

$$(V_1, V_2) = \frac{1}{F_\gamma(V_2, V_1)}$$

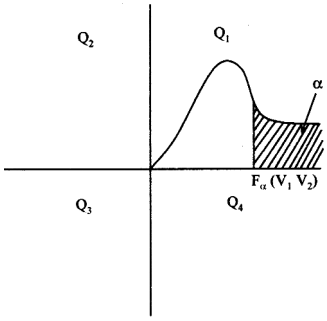


Fig.: F-Distribution Curve

Where,

$F_\alpha(V_1, V_2)$ = Value of F

So, α is at the right of $F_\alpha(V_1, V_2)$ under the F-distribution curve.

Applications

Following are the applications of F-distribution,

1. It is used for testing the equality of many population means.

2. It is used for comparing the sample variances.

3. It is used for performing analysis of variance.

4. It is used for testing the significance of regression equation.

5. It is used for determining whether the ratio incrementally changes from unity at any level chosen randomly.

Q2. Explain F-test for testing equality of two population variances.

Ans :

The objective of 'F-test' is to determine whether two independent estimates of the population variance differ significantly or whether the two samples may be regarded as drawn from the normal populations having the same variance, i.e.,

$$\sigma_A^2 = \sigma_B^2 = \sigma^2$$

This test is carried out by following ratio.

$$F = \frac{S_A^2}{S_B^2}$$

$$S_A^2 = \frac{\sum_{n=1}^{n_1} (x_a - \bar{x})^2}{n_1 - 1}$$

$$S_B^2 = \frac{\sum_{b=1}^{b_2} (y_b - \bar{y})^2}{n_2 - 1}$$

Where, n_i = 1st sample size;

n_2 = 2nd sample size.

PROBLEMS

1. Two samples are drawn from two normal populations. Test whether the two samples have the same variance at 5% level.

Sample 1	**60**	**65**	**71**	**74**	**76**	**82**	**85**	**87**		
Sample 2	**61**	**66**	**67**	**85**	**78**	**63**	**85**	**86**	**88**	**91**

Sol :

H_0 : The two populations have same variance.

Applying F-test, $F = \frac{s_1^2}{s_2^2} \sim F_{9,7,0.05}$

x	$(x - \bar{x})$	$(x - \bar{x})^2$	y	$(y - \bar{y})$	$(y - \bar{y})^2$
60	–15	225	61	–16	256
65	–10	100	66	–11	121
71	–4	16	67	–10	100
74	–1	1	85	8	64
76	1	1	78	1	1
82	7	49	63	–14	196
85	10	100	85	8	64
87	12	144	86	9	81
–	–	–	88	11	121
–	–	–	91	14	196
600		636	770		1200

$$\bar{x} = \frac{1}{8} \times 600 = 75$$

$$\bar{y} = \frac{1}{10} \times 770 = 77$$

$$s_1^2 = \frac{1}{n_1 - 1} \sum (x - \bar{x})^2 = \frac{1}{7} \times 636$$

$$s_1^2 = 90.857$$

$$s_2^2 = \frac{1}{n_2 - 1} \sum (y - \bar{y})^2 = \frac{1}{9} \times 1200$$

$$s_2^2 = 133.33$$

$$F = \frac{133.33}{90.857} \sim F_{9,7}$$

$$F = 1.468$$

Table value at 5% level,

$F_{9,7} = 3.68$

$F_{cal} = 1.468 < F_{9,7} = 3.68$

$\therefore$ Accept h_0

We conclude that, the population have same variance.

4.2 Analysis of Variance

4.2.1 Assumptions

Q3. Define is ANOVA? What are the assumptions and applications of ANOVA.

Ans : **(Imp.)**

Meaning

ANOVA is the acronym for Analysis of Variance. The variance test is also known as ANOVA. Analysis of variance is a statistical technique specially designed to test whether the means of more than two quantitative populations are equal i.e., to make inferences about whether those samples are drawn from the populations having the same mean.

The test is called 'F* test as it was developed by R.A Fisher in 1920's. The test is conducted in situations where we have three or more to consider at a time an alternative procedure (to t-test) needed for testing the hypothesis that all samples could likely be drawn from the same population.

Assumptions

1. Observations are independent

2. Each sample is drawn randomly from a normal population as the sample statistics reflect the characteristic of the population.

3. 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 fields mainly in their experiment design.

2. Anova helps to save time and money as several population means can be compared simultaneously.

3. Anova is used to test the linearity of the fitted regression line and correlation ratio, significance test statistic of anova = $F(r - 1, n - r)$.

4.2.2 Procedure

4.2.2.1 One way and two-way analysis of variance

Q4. Enlist the techniques of analysis ot variance. Explain about briefly one way ANOVA.

Ans : **(Imp.)**

Techniques of Analysis of Variance

There are two techniques of ANOVA. They are,

1. One way ANOVA

2. Two way ANOVA

For Answer Refer to Unit - IV, Q.No. 5

One Way ANOVA

In one way classification/single factor ANOVA, only one factor is considered and its effect on elementary units is observed, i.e., data are classified according to only one criterion. For instance, Yield of crop affected by type of seed only.

Procedure

The steps involved in computing one-way ANOVA are as follows,

Step-1:

Calculate the variance between samples,

$$SSB = \sum_{j=1}^{c} \frac{T_j^2}{n_j} - \frac{T^2}{N}$$

Step-2 :

Calculate the variance within samples,

$$SSW = \sum_{j=1}^{c}\sum_{i=1}^{n_j} X_{ij}^2 - \sum_{j=1}^{c} \frac{T_j^2}{n_j}$$

Step-3:

Calculate the F ratio as a ratio of mean squares between samples and mean square within samples.

Step-4:

Compare the calculated value of F ratio with table value of F for degrees of freedom for between and within samples.

Step-5:

If calculated F < table value of F, the difference is taken as insignificant i.e., due to change and we accept null hypothesis. If calculated value of F is equal to or more than table value, the difference is said to be significant.

ANOVA Table for One-way Classified Data

Source of variation	Sum of squares	d.f.	Mean sum of squares	Variable ratio
Treatment Ratio	S_t^2	k – 1	$\frac{S_t^2}{S_E^2} = F_{k-1}$	$\frac{S_t^2}{S_E^2} = F_{k-1}$, N –K
Error	S_E^2	N – K	$S_E^2 = \frac{S_E^2}{(N-K)}$	
Total	S_T^2	N – 1		

Q5. Write in detail about Two Way ANOVA.

Ans :

Meaning

Two way classification/two factor ANOVA is defined where two independent factors have an effect on the response variable of interest. For instance, Yield of crop affected by type of seed as well as type of fertilizer.

Procedure for Computation of Two-Way ANOVA

The steps involved in computation of two-way ANOVA one as follows,

Step-1:

Calculate the variance between columns,

$$SSC = \sum_{j=1}^{c} \frac{T_j^2}{n_j} - \frac{T^2}{N}$$

Step-2:

Calculate the variance between rows,

$$SSR = \sum_{i=1}^{r} \frac{T_i^2}{n_i} - \frac{T^2}{N}$$

Step-3:

Compute the total variance,

$$SST = \sum X_{ij}^2 - \frac{T^2}{N}$$

Step-4 :

Calculate the variance of residual or error,

SSE = TSS – (SSC + SSR)

Step-5 :

Divide the variances of between columns, between rows and residue by their respective degrees of freedom to get the mean squares.

Step-6:

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,

$$= \frac{\text{Mean square between rows}}{\text{Mean squares of residual}}$$

Step-7:

Compare F-ratio calculated with that of table value,

If F-ratio (calculated) < F-ratio (table), H_0 accepted,

If F-ratio (calculated) $\geq$ F-ratio (table), H_0 rejected.

H_0 accepted $\Rightarrow$ no significant differences

H_0 rejected $\Rightarrow$ significant differences

Two-Way ANOVA With Interaction

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 'S' (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,

$$\boxed{SST = SSA + SSB + SSAB + SSE}$$

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.

ANOVA Table for Two-way Classified Data with m-Observation Per Cell

Source of variation	d.f	S.S	M.S.S c	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			

Hypothesis Tests in Two-way ANOVA

Hypothesis tests in two-way ANOVA are as follows,

1. Factor A Test

Hypothesis is designed to determine whether there are any factor A main effects. Null Hypothesis is true if and only if there are no differences in means due to different treatments (population) of factor A.

2. Factor B Test

Hypothesis test is designed to detect the main effects of factor B. Null hypothesis is true if and only if there are no differences in means due to different treatments (populations) of factor B.

3. 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

2. To assess the significance of possible variation in performance in a certain test between the grammar schools of a city, a common test was given to a number of students taken at random. From the following results, make out an analysis of variance.

Schools			
A	B	C	D
8	12	18	13
10	11	12	9
12	9	16	12
8	14	6	16
7	4	8	15

To assess the significance of possible variationc in performance in a certain test between the grammar schools of a city is given below. From the following results make out an analysis of variance.

Schools			
A	B	C	D
8	12	18	13
10	11	12	9
12	9	16	12
8	14	6	16
7	4	8	15

Sol :

Step-1

Let the null hypothesis be that there is no significant variation in the performance of students in certain test i.e., $H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4$

Let the alternative hypothesis be that there is a significant variation in the performance of students in certain test i.e., $H_0 = \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

Schools							
School-A		School-B		School-C		School-D	
x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	x_4	x_4^2
8	64	12	144	18	324	13	169
10	100	11	121	12	144	9	81
12	144	9	81	16	256	12	144
8	64	14	196	6	36	16	256
7	49	4	16	8	64	15	225
45	421	50	558	60	824	65	875

Step-2: Calculation of Grand Total (GT)

GT = 45 + 50 + 60 + 65 = 220

Step-3: Calculation of Correction Factor

$$\text{Correction Factor} = \frac{(GT)^2}{N}$$

$n_1 = 5$, $n_2 = 5$, $n_3 = 5$ and $n_4 = 5$

$N = n_1 + n_2 + n_3 + n_4$

$N = 5 + 5 + 5 + 5 = 20$

$$\text{C.F.} \frac{(220)^2}{20} = 2{,}420$$

Step-4: Calculation of Total Sum of Squares (TSS)

$$TSS = \sum_i \sum_j X_{ij}^2 - \text{C.F.}$$

TSS = (421 + 558 + 824 + 875) – 2420

TSS = 2678 – 2420 = 258

Step-5: Calculation of Sum of Squares Between Samples (SSB)

$$SSB = \frac{\sum T_j^2}{n_j} \frac{GT^2}{N} = \left[\frac{45^2}{5} + \frac{50^2}{5} + \frac{60^2}{5} + \frac{65^2}{5}\right] - \frac{(220)^2}{20}$$

$$SSB = \left[\frac{2025}{5} + \frac{2500}{3} + \frac{3600}{5} + \frac{4225}{5}\right] - \frac{48400}{20}$$

SSB = [405 + 500 + 720 + 845] – 2420

SSB = 2470 – 2420 = 50

Step-6: Calculation of Sum of Squares within Samples

SSW = TSS – SSB

SSW = 258 – 50 = 208

Step-7: ANOVA Table

Sources of Vanation	Sum of Squares	Degrees of Freedom	Mean Squares	F-ratio
Between samples	50	(k –1) = (4 – 1) = 3	$\frac{50}{3}$ = 16.67	$\frac{16.67}{13}$ = 1.28
Within samples	208	(N–k) = (20–4) = 16	$\frac{208}{16}$ = 13	
Total	258	(N–1) = (20–1) = 19		

Step-8:

F-ratio calculated = 1.28

F-ratio from table when = 3 and V_2 = 16 is 3.24 at 5% level of significance.

Since, $F_{cal} < F_{table}$, we accept null hypothesis, i.e., there is no significant variation in the performance of students in a certain common test conducted between the different grammar schools of a city.

3. 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 : **(Imp.)**

Null Hypothesis

H_o : There is no significant difference between A,B,C and D.

Alternative Hypothesis

H_1 : There is a significant difference between A,B,C and D.

Soya bean

Block	Type A	Type B	Type C	Type D	Total
X	5	9	11	10	35
Y	4	7	8	10	29
Z	3	5	8	9	25
Total	12	21	27	29	GT = 89

Correction Factor (CF) = $\frac{(\text{Grand Total})^2}{N} = \frac{(89)^2}{12}$

= 660.08

Total Sum of Squares (TSS)

$$= \sum_i \sum_j X_{ij}^2 - \frac{(GT)^2}{N}$$

$= [(5)^2 + (9)^2 + (11)^2 + (10)^2 + (4)^2 + (7)^2 + (8)^2 + (10)^2 + (3)^2 + (5)^2 + (8)^2$
$+ (9)^21] - 660.08$

$= [25 + 81 + 121 + 100 + 16 + 49 + 64 + 100 + 9 + 25 + 64 + 81] - 660.8$

$= 735 - 660.08$

$\therefore$ TSS = 74.92

Sum of Squares Between Soya Bean (Columns)

$$SSB = \sum_j \frac{T_j^2}{n_j} - \frac{(GT)^2}{N}$$

$$= \frac{(12)^2}{3} + \frac{(21)^2}{3} + \frac{(27)^2}{3} + \frac{(29)^2}{3} - 660.08$$

$= [48 + 147 + 243 + 280.33] - 660.08$

$= 718.33 - 660.08$

$\therefore$ SSB = 58.25

Degree of freedom (r) = (K – 1)

= (4 – 1)

= 3

Mean sum of squares between the soya beans = $\frac{58.25}{3}$ = 19.42

Sum of Squares within Blocks (SSW)

SSW = TSS – SSB

= 74.92 – 58.25

= 16.67

Mean sum of squares within the blocks

$$= \frac{16.67}{12-4} = \frac{16.67}{8} = 2.08$$

ANOVA TABLE

Sources of variation	Sum of squares	Degrees of Freedom	Means squares
Between soya bean type	58.25	(k – 1) = (4 – 1) = 3	19.42
Within blocks	16.67	(n – k) = (12 – 4) = 8	2.08
Total		(n – 1) = (12 – 1) = 11	

$$\text{F-Ratio} = \frac{\text{Mean square between soya bean type}}{\text{Mean square within blocks}}$$

$$= \frac{19.42}{2.08} = 9.34$$

[Note: Assuming level of significance as 5%]

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

Decisions

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.

4. Four technicians analyzed three samples each for the moisture content in the sample. The results are given below:

Samples	Technicians			
	A	B	C	D
X	9	12	10	11
Y	12	11	15	12
Z	9	10	12	14

Analyze the data and comment. Use 5% significance level.

Sol :

A two-way ANOVA technique will enable us to analyse the data and comment.

Here, the two factors are, Technicians (T) and Samples (S).

Step-1: Null Hypothesis

$H_{OT} = \mu_A = \mu_B = \mu_C = \mu_D$ i.e., the mean moisture content is same according to all technicians.

$H_{OS} = \mu_X = \mu_Y = \mu_Y = \mu_Z$ i.e., the mean moisture content is same in all the samples.

Alternative Hypothesis

H_{IT} = Atleast two of μ_A, μ_B, μ_C and μ_D are different.

H_{IS} = Atleast two of μ_X, μ_Y, μ_Z are different.

Samples	Technicians					
	A	B	C	D	Row Totals(R)	R^2
X	9	12	10	11	R_1 = 42	R_1^2 = 1764
Y	12	11	15	12	R_2 = 50	R_2^2 = 2500
Z	9	10	12	14	R_3 = 45	R_3^2 = 2025
Column Totals (C)	30	33	37	37	137	
C^2	900	1089	1369	1369		

Step-2: Calculation Sum of Items of all Samples

$GT = \Sigma\Sigma x_{ij} = \Sigma_i R_i = \Sigma_j C_j$

GT = 30 + 33 + 37 + 37 = 137

Step-3: Calculation of Correction Factor (C.F)

Here, GT = 137, N = 12,

$$C.F = \frac{(GT)^2}{N} = \frac{(137)^2}{12} = \frac{18769}{12} = 1564.08$$

Step-4: Calculation of Total Sum of Squares (TSS)

$TSS = \Sigma_i \Sigma_j X_{ij}^2 - C.F$

$TSS = (9^2 + 12^2 + 10^2 + 11^2 + 12^2 + 11^2 + 15^2 + 12^2 + 9^2 + 10^2 + 12^2 + 14^2) - 1564.08$

TSS = (81 + 144 + 100 + 121 + 144 + 121 + 225 + 144 + 81 + 100 + 144 + 196) – 1564.08

TSS = 1601 – 1564.08 = 36.92

Step-5: Calculation of Sum of Squares Between Rows (i.e., Between Samples)

$$SSR = \frac{R_1^2}{n} + \frac{R_2^2}{n} + \frac{R_3^2}{n} + ... + \frac{R_k^2}{n} - C.F.$$

$$SSR = \left[\frac{(42)^2}{4} + \frac{(50)^2}{4} + \frac{(45)^2}{4}\right] - 1564.08$$

SSR = (441 + 625 + 506.25) – 1564.08

SSR = (441 + 625 + 506.25) – 1564.08

SSR = 1572.25 – 1564.08

SSR = 8.17

Step-6: Calculation of Sum of Squares Between Columns (i.e., Between Technicians)

$$SSC = \frac{C_1^2}{n} + \frac{C_2^2}{n} + \frac{C_3^2}{n} + ... + \frac{C_k^2}{n} - C.F$$

$$SSC = \left(\frac{30^2}{3} + \frac{33^2}{3} + \frac{37^2}{3} + \frac{37^2}{3} \right) - 1564.08$$

SSC = (300 + 363 + 456.33 + 456.33) – 1564.08

SSC = 1575.66 – 1564.08 = 11.58

Step-7: Calculation of Sum of Squares of Residual or Error (SSE)

SSE = TSS – SSR – SSC

= 36.92 – 8.17 – 11.58 = 17.17

ANOVA Table

Sources of Vanation	Sum of Squares	Degrees of Freedom	Mean Sum of Squares	F-ratio	F-ratio
Between Technicians	11.58	(c – 1) = (4 – 1) = 3	$\frac{11.58}{3} = 3.86$	$\frac{3.86}{2.86} = 1.35$	4.76
Between samples	8.17	(r – k) = (3 – 1) = 2	$\frac{8.17}{2} = 4.09$	$\frac{4.09}{2.86} = 1.43$	5.14
Residual Error	17.17	(c – 1) = (r – 1) = 3 × 2 = 6	$\frac{17.17}{6} = 2.86$		
Total	36.92				

Step-8:

Table values of F-ratio at 5% level of significance.

$F_{(6,3)} = 4.76$

$F_{(6,2)} = 5.14$

(i) Calculated $F_{(6,3)}$ < Table $F_{(6,3)}$

Hence null hypothesis is accepted i.e., there is no significant difference between technicians opinion about mean moisture content in the sample.

(ii) Calculated $F_{(6,2)}$ < Table $F_{(6,2)}$

Hence, null hypothesis is accepted i.e., there is no significant difference between samples. Mean moisture content is same in all the samples.

4.3 Association of Attributes

4.3.1 Meaning

Q6. Explain briefly about Association of Attributes.

Ans : **(Imp.)**

Meaning

Association in statistical language is quiet different from the common meaning of association. Commonly, if two attributes A and B appear together number of times then they can be said to be as associated.

Definition

According to Yule and Kendall, "In Statistics A and B are associated only if they appear together in a greater number of cases than is to be expected, if they are independent."

Methods used to measure the association of attributes refer to those techniques, which are used to measure the relationship between two such phenomena, whose size cannot be measured and where we can only find the presence or absence of an attribute.

Statistics deals with quantitative phenomenon only. However, the quantitative character may arise in any of the following two ways:

4.4 Correlation

Q7. Define correlation. Explain the significance of correlation.

Ans : **(Imp.)**

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".

(iii) **According to Simpson's and Kafka,** "Correlation analysis deals with the association between two or more variables.

Significance

1. Correlation 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. 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.

3. The coefficient of correlation is a relative measure and we can compare the relationship between variables, which are expressed in different units.

4. Correlations are useful in the areas of health care 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.

5. Sampling error can also be calculated.

6. Correlation is the basis for the concept of regression and ratio of variation.

7. The decision making is heavily felicitated by reducing the range of uncertainty and hence empowering the predictions.

Q8. Explain the scope of correlation analysis.

Ans :

Scope

(i) One of the variable may be affecting the other

A correlation coefficient calculated from the data on quantity demanded and corresponding price of tea would only reveal that the degree of association between them is very high. It will not give us any idea about whether price is affecting demand of tea or vice-versa. In order to know this, we need to have some additional information apart from the study of correlation.

(ii) The two variables may act upon each other

Cause and effect relation exists in this case also but it may be very difficult to find out which of the two variables is independent.

(iii) The two variables may be acted upon by the outside influences:

In this case we might get a high value of correlation between the two variables, however, apparently no cause and effect type relation seems to exist between them.

(iv) A high value of the correlation coefficient may be obtained due to sheer coincidence (or pure chance)

This is another situation of spurious correlation. Given the data on any two variables, one may obtain a high value of correlation coefficient when in fact they do not have any relationship.

Q9. What are the Properties of Correlation?

Ans :

Properties

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.

4.5 Distinction between Correlation and Association

Q10. Distinction between Correlation and Association of attributes.

Ans :

S.No.	Correlation coefficient	S.No.	Association of attributes
1.	The degree of the correlation that exists between the variables is correlation corfficient (y).	1.	Association of attributes measures the degree of relationship between two attributes (Q).
2.	eg: (a) Demand and supply. (b) Price and demand.	2.	eg: (a) Sex and literacy. (b) Beauty and intelligence.
3.	$y = \pm 1$ is the range	3.	$Q \pm 1$ is the range.

4.6 Methods of Studying Association - Interpretation of Results

Q11. What are the various methods of studying association? Explain briefly about comparison of observed and expected frequencies methods.

Ans :

In order to ascertain whether two attributes are associated or not the following methods may be used:

1. Comparison of observed and expected frequencies methods
2. Proportion method
3. Yule's Coefficient of Association
4. Coefficient of Colligation
5. Coefficient of Contingency

1. Comparison of Observed and Expected Frequencies Methods

When this method is applied, the actual observation is compared with the expectation. If the actual observation is equal to the expectation, the attributes are said to be independent; if actual observation is more than the expectation, the attributes arc said to be positively associated and if the actual observation is less than the expectation, the attributes are said to be negatively associated.

Symbolically, attributes A and B are :

(i) Independent if $(AB) = \frac{(A)\times(B)}{N}$ (expectation) ; (actual observation)

(ii) Positively associated if $(AB) > \frac{(A)\times(B)}{N}$ (expectation) ; and (actual observation)

(iii) Negatively associated if $(AB) < \frac{(A)\times(B)}{N}$ (expectation) ; (actual observation)

The same is true for attributes a and B; α and β and A and b. The attributes a and b shall be called:

(i) Independent, if $(\alpha\beta) = \frac{(\alpha)\times(\beta)}{N}$;

(ii) Positively associated, if $(\alpha\beta) > \frac{(\alpha)\times(\beta)}{N}$; and

(iii) Negatively associated, if $(\alpha\beta) > \frac{(\alpha)\times(\beta)}{N}$.

2. Proportion method

For Answer Refer to Unit - IV, Q.No. 12

3. Yule's Coefficient of Association

For Answer Refer to Unit - IV, Q.No. 13

4. Coefficient of Colligation

For Answer Refer to Unit - IV, Q.No. 14

5. Coefficient of Contingency

For Answer Refer to Unit - IV, Q.No. 15

PROBLEMS

5. Show whether A and B are independent, positively associated or negatively associated in each of the following cases:

(i) N = 1000; (A) = 450; (B) = 600; (AB) = 340

(ii) (A) = 480; (AB) = 290; (α) = 585; (αB) = 383

(iii) N = 1000; (A) = 500; (B) = 400; (AB) = 200

Sol:

We have given

(i) $\frac{(A)(B)}{N} = \frac{450\times600}{1,000} = 270 = (AB)_0$

Thus, $(AB) = 340 > \frac{(A)(B)}{N}$

Since $(AB) > (AB)_0$ hence they are positively associated.

(ii) $\because (B) = (AB) + (\alpha B) = 290 + 383 = 673$

$N = (A) + (\alpha) = 480 + 585 = 1065$

$\therefore \frac{(A)(B)}{N} = \frac{480\times673}{1065} = 303.32\ (AB)_0$

Thus, $(AB) = 290 < 303.32$

$\because (AB) < (AB)_0$

$\therefore$ A and B negatively associated

(iii) $\frac{(A)(B)}{N} = \frac{500 \times 400}{1000} = 200 = (AB)_0$

Thus, we find $(AB) = (AB)_0$

Hence, A and B are independent, i.e. $\delta = 0$.

6. Find if A and B are independent, positively associated or negatively associated in each of the following cases:

(i) N = 100; (A) = 47; (B) = 62 and (AB) = 32

(ii) (A) = 495; (AB) = 292; (α) = 572 and ($\alpha\beta$) = 380

(iii) (AB) = 2560; (αB) = 7680; (Aβ) = 480 and ($\alpha\beta$) = 1440

Sol :

(i) N = 100; (A) = 47; (B) = 62 and (AB) = 32

$$(AB)_0 = \frac{(A)(B)}{N} = \frac{47 \times 62}{100} = 29.14$$

$\therefore$ $(AB) = 32 > (AB)0$

$\therefore$ A and B are positively related.

(ii) (A) = 495; (AB) = 292; (α) = 572 and ($\alpha\beta$) = 380

We have

$N = (A) + (\alpha) = 495 + 572 = 1067$

$(B) = (AB) + (\alpha B) = 292 + 380 = 672$

$\therefore$ $(AB)_0 = \frac{(A)(B)}{N} = \frac{495 \times 672}{1067} = 31.75$

$\because$ $(AB) = 292 < (AB)_0$

$\therefore$ A and B are negatively related.

(iii) (AB) = 2560; (αB) = 7680; (Aβ) = 480 and ($\alpha\beta$) = 1440

$(A) = (AB) + (A\beta) = 2560 + 480 = 3040$

$(B) = (AB) + (\beta B) = 2560 + 7680 = 10240$

$N = (AB) + (A\beta) + (\alpha B) + (\alpha\beta)$

$= 2560 + 480 + 7680 + 1440$

$= 12160$

$\therefore$ $(AB)_0 = \frac{(A)(B)}{N} = \frac{3040 \times 10240}{12160} = 2560$

$\because$ $(AB) = 2560 = (AB)_0$

$\therefore$ A and B are independent.

$\therefore$ i.e., $\delta = 0$.

12. Explain briefly about proportion method.

Ans :

If there is no relationship of any kind between two attributes A and S-we expect to find the same proportion of A's amongst the B's as amongst the P's. Thus, if a coin is tossed we expect the same proportion of heads irrespective of whether the coin is tossed by the right hand or the left hand.

Symbolically, two attributes may be termed :

(i) Independent if $\frac{(AB)}{B} = \frac{(A\beta)}{(\beta)}$

(ii) Positively associated if $\frac{(AB)}{B} > \frac{(A\beta)}{(\beta)}$

(iii) Negatively associated if $\frac{(AB)}{B} < \frac{(A\beta)}{(\beta)}$

If the relation (i) holds good the corresponding relations

$$\frac{(\alpha B)}{(B)} = \frac{(\alpha\beta)}{(\beta)}; \frac{(AB)}{A} = \frac{(\alpha B)}{(\alpha)}; \frac{(AB)}{(A)} = \frac{(\alpha\beta)}{(\alpha)}$$

must also hold true.

PROBLEMS

7. Find if A and B are independent, positively associated or negatively associated from the data given below :

(A) = 470, (B) = 620, (AB) = 32, N = 1000

Sol :

Attributes A and B shall be called independent if

$$(AB) = \frac{(A)\times(B)}{N}$$

(AB) = 320, (A) = 470, (B) = 620, N = 1000

Expectation of (AB) = $\frac{470\times620}{1000}$ = 291.4

Since (AB) actual observation (320) is mote that he expectation (291.4) attributes A and B are positively associated.

8. In a population of 500 students the number of married is 200. Out of 150 students who failed 60 belonged to the married group. It is required to find out whether the attributes marriage and failure are independent, positively associated or negatively associated.

Sol :

Let *A* denote married students.

$\therefore$ α represents unmarried students.

Let B denote number of failures.

$\therefore$ β would denote non-failures.

We are given the total number of students, i.e., N = 500

(A) = 200 , (B) = 150 and (AB), i.e., the number of married students who failed, = 60.

Attributes A and B shall be called independent if $\frac{(AB)}{(A)} = \frac{(\alpha B)}{(\alpha)}$

In other words, if the proportion of married students who failee Hs the same as the proportion of unmarried students who failed we say that the attributes, marriage and failure, are independent.

Proportion of unmarried students who failed : i.e., = $\frac{(AB)}{(A)} = \frac{60}{200}$ = 0.3 or 30%

Proportion of unmarried students who failed :

i.e., $\frac{(\alpha B)}{(\alpha)} = \frac{90}{300}$ = 0.3 or 30% $\left\{\begin{matrix}(\alpha B = (B) = (AB) \text{ i.e., } 150 - 60 = 90 \\ (\alpha) = N - (A), \text{ i.e., } 500 - 200 = 300\end{matrix}\right\}$

Since the two proportions are the same we conclude that the attributes, marriage and failure, are independent.

Q13. Explain Yule's coefficient of association.

Ans : **(Imp.)**

The most popular method of studying association is the Yule's Coefficient because here not only we can determine the nature of association, *i.e.,* whether the attributes are positively associated, negatively associated or independent, but also the degree or extent to which the two attributes are associated. The Yule's Coefficient is denoted by the symbol Q and is obtained by applying the following formula:

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

The value of this coefficient lies between ± 1. When the value of Q is + 1 there is perfect positive association between the attributes. When Q is – 1 there is perfect negative association (or perfect dissociation) between the attributes and when the value Q is zero the two attributes are independent.

The coefficient of association can be used to compare the intensity of association between two attributes with the intensity of association between two other attributes.

PROBLEMS

9. Investigate the association between eye colour of husbands and eye colour of wives from the data given below :

Husbands with light eyes and wives with light eyes = 309

Husbands with light eyes and wives with not-light eyes = 214

Husbands with not-light eyes and wives with light eyes = 132

Husbands with not-light eyes and wives with not-light eyes = 119

Sol :

Since we have to find out the association between eye colour of husband and that of wife, one attribute we would take as A and another as B.

Let A denote husbands with light eyes.

$\therefore \alpha$ would denote husbands with not-light eyes.

Let B denote wives with light eyes

$\therefore \beta$ would denote wives with not-light eyes.

The given data in terms of these symbols is $(AB) = 309$, $(A\beta) = 214$, $(\alpha B) = 132$, $(\alpha\beta) = 119$.

Applying Yule's method : $Q = \dfrac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$

Substituting the above values in the formula

$$Q = \frac{(309 \times 119) - (214 \times 132)}{(309 \times 119) + (214 \times 132)} = \frac{8523}{65019} = 0.131$$

Thus, there is a very little association between the eye colour of husband and wife.

10. Eighty-eight residents of an Indian city, who were interviewed during a sample survey, are classified below according to their smoking and tea drinking habits. Calculate Yule's Coefficient of Association and comment on its value.

Particulars	Smokers	Non-Smokers
Tea Drinkers B	40	33
Non-tea Drinkers β	3	12

Sol:

Let *A* denote smokers.

$\therefore \alpha$ would denote non-smokers.

Let B denote tea drinkers.

$\therefore \beta$ would denote non-tea drinkers.

The given data in terms of these symbols are :

(AB), i.e., number of smokers and tea drinkers = 40

(Aβ), i.e., number of smokers and non-tea drinkers = 3

(αβ), i.e., number of tea drinkers and non-smokers = 33

(αβ), i.e., number of non-smokers and non-tea drinkers = 12.

Applying Yule s method : $Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$

Substituting the values of (AB), (Aβ), (αB) and (αβ) in this formula

$$Q = \frac{(40 \times 12) - (3 \times 33)}{(40 \times 12) + (3 \times 33)} = \frac{480 - 99}{480 + 99} = \frac{381}{579} = 0.658$$

This shows that the attributes tea drinking and smoking are positively associated.

11. The following table is reproduced from a memoir written by Karl Pearson.

Eye colour in father	Eye colour in son	
	Not light	**Light**
Not light	**230**	**148**
Light	**151**	**471**

Discuss whether the colour of the son's eye is associated with that of father.

Sol :

Let A represents the light eye colour of father an B represents the light eyecolour of son. Then α represents not light eye colour of father and b represents not light eye colour of son. Then the given data is,

$(\alpha\beta) = 230, (\alpha B) = 148$

$(A\beta) = 151, (AB) = 471$

Coefficient of Association is

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(471 \times 230) - (151 \times 148)}{(471 \times 230) + (151 \times 148)}$$

$$= \frac{85982}{130678} = +\ 0.657$$

This shows that there is fairly high degree of positive association between eye colour of father and son.

Q14. Explain coefficient of colligation association.

Ans :

Yule has computed another coefficient called the coefficient of 'colligation'. It is denoted by the symbol γ and is obtained by applying the following formula :

$$\gamma = \frac{1-\sqrt{\frac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}}{1+\sqrt{\frac{(A\beta)\times(\alpha B)}{(AB)\times(\alpha\beta)}}}$$

From this coefficient we can obtain Yule's Coefficient of Association, i.e., Q as follows:

$$Q = \frac{2\gamma}{1+\gamma^2}$$

It should be noted that though γ and Q serve the same purpose, These coefficients are not directly comparable with each other. Further, in practice Q is more popularly used than γ as a measure of association.

PROBLEMS

12. Given

(AB) = 35 ($\alpha\beta$) = 7

(Aβ) = 8 (αB) = 6

Calculate the coefficient of colligation.

Sol :

Coefficient of colligation

$$\gamma = \frac{1-\sqrt{\frac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}}{1+\sqrt{\frac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}} = \frac{1-\sqrt{\frac{8\times6}{35\times7}}}{1+\sqrt{\frac{8\times6}{35\times7}}}$$

$$= \frac{1-\sqrt{0.196}}{1+\sqrt{0.196}} = \frac{1-0.44}{1+0.44} = 0.39$$

Thus, two attributes A and B are positively associated.

Q15. Explain a contingency table with an example?

(OR)

What is a contingency table? Explain the method of its construction.

(OR)

What do you mean by the coefficient of contingency? When does it assume maximum value?

Ans : **(Imp.)**

Qualitative data are often classified into more than two classes, *i.e.*, attribute A may be classified not as 'A' and 'not A', but as A_1, A_2, A_3, etc. Similarly, another attribute B may be subdivided into B_1, B_2, B_3, etc. The frequencies falling within the different classes may be arranged in the form of a Contingency Table as follows:

B / A	B_1	B_2	B_3	...	B_n	Total
A_1	$(A_1 B_1)$	$(A_1 B_2)$	$(A_1 B_3)$	...	$(A_1 B_n)$	(A_1)
A_2	$(A_2 B_2)$	$(A_2 B_2)$	$(A_2 B_3)$	...	$(A_2 B_n)$	(A_2)
A_3	$(A_3 B_1)$	$(A_3 B_3)$	$(A_3 B_3)$	...	$(A_3 B_n)$	(A_3)
...	...	...	...	...	...	...
A_n	$(A_n B_1)$	$(A_n B_2)$	$(A_n B_3)$	...	$(A_n B_n)$	(A_n)
Total	(B_1)	(B_2)	(B_3)	...	(B_n)	N

For determining the degree of association between A's and B's on the whole the coefficient of mean square contingency as given by Pearson may be used. The coefficient of mean square contingency is denoted by the symbol *i* and obtained by applying the following formula :

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

While finding out the value of i we proceed on the assumption of null hypothesis, i.e., the two attributes are independent and exhibit no association.

For calculation of i we have to determine the value of χ^2 (pronounced as chi-square)*.

The steps in calculating the value of χ^2

(i) Find the expected or independent frequency for each cell. Thus, for cell $(A_1 B_1)$ the expectation is $\frac{(A_1) \times (B_1)}{N}$.

(ii) Obtain the difference between the observed (actual) and expected frequencies in each cell, *i.e.,* find (O – E).

(iii) Square (O – E) and divide the figure by E, the expected frequency for each cell.

(iv) Add up the figures obtained in step (iii).

(v) This would give the value of χ^2 Thus, $\chi^2 = \Sigma \frac{(O - E)^2}{E}$.

Once the value of χ^2 is obtained it is easy to determine the value of i.

4.7 Chi Square Test

4.7.1 Definition

Q16. What is Chi-Square (χ^2)? Explain the applications test statistic and assumptions.

Ans : **(Imp.)**

Meaning

The square of a standard normal variable is called a chi-square (χ^2) which variate with one degree of freedom (d.f). It is expressed as,

$$\chi^2 = \left(\frac{x-\mu}{\sigma}\right)^2$$

Where 'χ' is a random variable lollowing normal distribution with mean 'μ' is an standard deviation a. If χ_1, χ_2.....χ_V are independent random variables following normal distribution with means μ_1, μ_2,...,μ_V and standard deviations σ_1, χ_2,...,χ_V then,

$$\chi^2 = \sum\left(\frac{X-\mu}{\sigma}\right)^2$$

χ^2 is the sum of square of V independent standard normal variants following chi-square distribution. χ^2 is a family of distributions one for each value of V.

Applications

Chi-square distribution has a number of applications,

1. Chi-square test of goodness of fit.

2. Chi-square test for independence of attributes.

3. To test if population has a specified value of variance.

4. To test equality of several population proportions.

Chi-square Test Statistic

Chi-square is an important non-parametric test and does not have any prerequisites (no assumptions) in respect of the type of populations.

χ^2 describes the magnitude of discrepancy between theory and observations. Whether it can be attributed to chance or it is due to inadequacy of theory to fit the observed facts. If x^2 is zero, observed and expected frequencies are equal. The greater x^2 value more would be the discrepancy between observed and expected frequencies.

$$\chi^2 = \sum\frac{(O-E)^2}{e}$$

Where , O = Observed frequency

E = Expected frequency

Calculated value of χ^2 is compared with table value of x^2 for given degrees of freedom at a specified level of significance.

(a) If χ^2 (calculated value) > χ^2 (table value) the difference between theory and observation is considered to be significant (fluctuations of sampling not responsible).

(b) If χ^2 (calculated value) < χ^2 (table value) the difference between theory and observation is not considered significant (fluctuations in sampling responsible).

Degrees of freedom have restriction of the total sample observations.

Assumptions

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 in original units to make comparison easy.

4. Sample size should have atleast 50 observations.

4.7.2 Conditions for applying Chi square test

Q17. What are the basic conditions for the application of Chi Square Test?

Ans :

The following conditions should be satisfied before applying the χ^2 test :

1. In the first place N must be reasonably large to ensure the similarity between theoretically correct distribution and our sampling distribution of χ^2, the chi-square statistic. It is difficult to say what constitutes largeness, but as a general rule χ^2 test should not be used when N is less than 50, however few the cells.

2. No theoretical cell frequency should be small when the expected frequencies are too small, the value of χ^2 will be overestimated and will result in too many rejections of the null hypothesis. To avoid making incorrect inferences, a general rule is followed that expected frequency of less than 5 in one cell of a contingency table is too small to use. When the table contains more than one cell with an expected frequency of less than 5 we "pool" the frequencies which are less than 5 with the preceding or succeeding frequency so that the resulting sum is 5 or more. However, in doing so, we reduce the number of categories of data and will gain less information from contingency table.

3. The constraints on the cell frequencies if any should be linear, i.e., they should not involve square and higher powers of the frequencies such as $\Sigma O = \Sigma E = N$.

4.7.3 Yates's Correction

Q18. Explain briefly about Yates's correction.

Ans : **(Imp.)**

The Yates' corrections, also called Yates' corrections for continuity, are introduced because theoretical chi-square distribution is continuous whereas the tabulated values are based on the distribution of discrete χ^2 statistic. The corrections has the effect of reducing the calculated value of χ^2 as continuous compared to the corresponding value without correction.

In a special case of 2 x 2 contingency table the approximation may be improved, and bias arising out of the use of small theoretical frequencies may be reduced, by means of a correction proposed by F. Yates in 1934. The correction involves the reduction of the deviation of observed from theoretical frequencies which of course reduces the value of χ^2. The working rule for the application of the correction is : adjust the observed frequency in each cell of the 2×2 table in such a way as to reduce the absolute deviation of the $\frac{1}{2}$ observed from the theoretical frequency for that cell by adjustments for all the cells are to be made without changing the marginal totals. This operation will increase f_0, that is observed frequency, by $\frac{1}{2}$ in each of two cells, and will reduce f_0 by $\frac{1}{2}$ in each of two cells.

Another method of adjustment which gives the same result as the above procedure is :

$$\chi^2 \text{ (corrected } \Sigma - \frac{[O_1 - E_1 - 0.5]^2}{E_1} + \frac{[O_2 - E_2 - 0.5]^2}{E_2} + \frac{[(O_k - E_k - 0.5]^2}{E_k} \text{ or}$$

$$\chi^2 = S \frac{[|O - E| - 0.5]^2}{E}$$

In general, correction is made only when the number of degrees of freedom or v = 1 and N is small. For large samples this yields practically the same result as the uncorrected χ^2. For small samples where each expected frequency is between 5 and 10, it is perhaps best to compare both the corrected and uncorrected values of χ^2. If both values lead to the same conclusion regarding a hypothesis, such as acceptance or rejection at 0.05 level, difficulties are rarely encountered. If they lead to different conclusions, one can either resort to increasing sample sizes or if this proves impractical, one can employ exact methods of probability involving the multinomial distribution.

4.7.4 Uses of Chi Square Test, Limitations of Chi Square Test

Q19. What are the uses of Chi Square Test?

(OR)

Give various uses of Chi Square Test.

(OR)

Chi Square Test is a test of Independence, Homogeneity and Goodness of fit. Discuss briefly.

(OR)

Explain two uses of Chi Square distribution in test of significance.

(OR)

What are the Limitations of Chi Square Test?

Ans : **(Imp.)**

Uses

The χ^2 test is one of the most popular statistical inference procedures today. It is applicable to a very large number of problems in practice which can be summed up under the following heads:

i) **χ^2 test as a test of independence:** With the help of χ^2 test we can find out whether two or more attributes are associated or not.

ii) **χ^2 test as a test of goodness of fit:** χ^2 test is very popularly known as test of goodness of fit for the reason that is enables us to ascertain how appropriately the theoretical distributions such as Binomial, Poisson, Normal, *etc*., fit empirical distributions, *i.e*., those obtained from sample data. When an ideal frequency curve whether normal or some other type is fitted to the data, we are interested in finding out how well this curve its with the observed facts.

iii) **χ^2 test as a test of homogeneity:** The χ^2 test of homogeneity is an extension of the chi-square test of independence. Tests of homogeneity are designed to determine whether two or more independent random samples are drawn from the same population or from different populations. Instead of one sample as we use with independence problem we shall now have two or more samples.

Limitations

χ^2 test is very widely used in practice. However, in order to avoid the misapplication of the test its following limitations should be kept in mind:

i) Frequencies of non-occurrence should not be omitted for binomial or multinomial events.

ii) The formula presented for χ^2 statistics in terms of frequencies. Hence and attempt should not be made to compute on the basis of proportions or another derived measures.

4.7.5 Chi Square Test for Testing the Independence of Attribute

Q20. Explain briefly about Chi square test for testing the independence of Attribute.

Ans : **(Imp.)**

χ^2 test facilitates us to explain whether or not two attributes are associated. χ^2 is not a measure of the degree of relationship or the form of relationship between two attributes, but is simply a technique of judging the significance of such association or relationship between two attributes.

If the calculated value of χ^2 is less than table value, then null hypothesis is accepted which means the two attributes are independent and not associated. Otherwise, null hypothesis is not accepted which means the two attributes are associated and the association is not because of some chance but it exists in reality.

Procedure of Test for Independence of Attributes

Following are the steps involved in determining chi-square using test for independence of attributes.

1. Expected frequencies are calculated for each cell using the following equation.

$$E = \frac{RT \times CT}{G}$$

Where E = Expected frequency

RT = Row total for the row of cell

CT = Column total for the column of cell

N = Total number of observations

2. Compute differences between observed and expected frequencies and find their squares i.e., $(O - E)^2$.

3. Divide the value of $(O - E)^2$ with respective expected frequency and find the sum $\Sigma[(O - E)^2/E]$.

4. Find the degree of freedom for 2×2 table i.e., $(2 - 1)(2 - 1) = 1$. At 1 d.f find the table value of chi-square at given level of significance.

5. Compare table value with computed value of chi-square at specified level of significance. If $\chi^2_{cal)} > \chi^2_{table}$ $\Rightarrow$ two attributes are dependent. $\chi^2_{cal)} > \chi^2_{table}$ $\Rightarrow$ two attributes are independent.

PROBLEMS

13. Following is the classification of 100 students according to their sex and height. Test whether the height of the students is dependent upon the sex.

	Height	
Sex	**Tall**	**Short**
Male	**30**	**40**
Female	**20**	**10**

Sol :

Null Hypothesis

H_0 : Sex and height of students are independent i.e., they are not associated with each other.

Alternate Hypothesis

H_a : The height of students are dependent upon the sex.

$$\text{Expected frequency} = \frac{\text{Row Total} \times \text{Colum Total}}{\text{Grand Total}}$$

$$E_{ij} = \frac{RT_i \times CT_j}{GT}$$

$RT_1 = 30 + 40 = 70$

$RT_2 = 20 + 10 = 30$

$CT_1 = 30 + 20 = 50$

$CT_2 = 40 + 10 = 50$

Grand Total (or) GT = 70 + 30 = 100

or

= 50 + 50 = 100

$$E_{ij} = \frac{RT_i \times CT_j}{GT}$$

$$E_{11} = \frac{70 \times 50}{100} = 35$$

$$E_{12} = \frac{70 \times 50}{100} = 35$$

$$E_{21} = \frac{30 \times 50}{100} = 15$$

$$E_{22} = \frac{30 \times 50}{100} = 15$$

Chi-square Table

Cell	O	E	O – E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
E_{11}	30	35	– 5	25	0.714
E_{12}	40	35	5	25	0.714
E_{21}	20	15	5	25	1.67
E_{22}	10	15	– 5	25	1.67
			0		**Total = 4.768**

$$\chi^2 = \Sigma\left[\frac{(O-E)}{E}\right]$$

Where,

O = Observed value of i^{th} row and j^{th} column

E = Expected value of i^{th} row and j^{th} column

From the above table (chi-square table) $\chi^2 = 4.768$

Degree of freedom,

$(r - 1)(c - 1)$

$(2 - 1)(2 - 1)$

$d.f = 1 \times 1$

Level of significance,

$\alpha = 5\%$ (assumed)

χ^2 table for d.f = 1

$\alpha = 0.05$ is 3.841

Since, $\chi^2_{cal} \geq \chi^2_{table}$ (4-768 > 3.841)

Decision

Null hypothesis is rejected which means that the sex and height is associated with each other.

14. Four machines A, B, C, D are used to manufacture parts which are classified as first, second and third grade. Test whether the quality of the product is independent of the machines. The data as under,

Grade	A	B	C	D
First	620	750	400	530
Second	130	200	140	130
Third	50	50	60	40

Sol : **(Imp.)**

Null Hypothesis

H_0 : Quality of product and machines are independent i.e., they are not associated with each other.

Alternate Hypothesis

H_a : Quality of product and machines are associated.

$$\text{Expected frequency} = \frac{\text{Row Total} \times \text{Colum Total}}{\text{Grand Total}}$$

First Row Total = 620 + 750 + 400 + 530 = 2300

Second Row Total = 130 + 200 + 140 + 130 = 600

Third Row Total = 50 + 50 + 60 + 40 = 200

A Column Total = 620 + 130 + 50 = 800

B Column Total = 750 + 200 + 50 = 1000

C Column Total = 400 + 140 + 60 = 600

D Column Total = 530 + 130 + 40 = 700

Grand Total = 2300 + 600 + 200 = 3100

(or)

= 800 + 1000 + 600 + 700 = 3100

$$E_{11} = \frac{RT_1 \times CT_1}{GT} = \frac{2300 \times 800}{3100} = 593.5$$

$$E_{12} = \frac{RT_1 \times CT_2}{GT} = \frac{2300 \times 1000}{3100} = 741.9$$

$$E_{13} = \frac{RT_1 \times CT_3}{GT} = \frac{2300 \times 600}{3100} = 445.2$$

$$E_{14} = \frac{RT_1 \times CT_3}{GT} = \frac{2300 \times 700}{3100} = 519.4$$

$$E_{21} = \frac{RT_2 \times CT_1}{GT} = \frac{600 \times 800}{3100} = 154.8$$

$$E_{22} = \frac{RT_2 \times CT_2}{GT} = \frac{600 \times 600}{3100} = 193.5$$

$$E_{23} = \frac{RT_2 \times CT_2}{GT} = \frac{600 \times 600}{3100} = 116.1$$

$$E_{24} = \frac{RT_2 \times CT_2}{GT} = \frac{600 \times 700}{3100} = 135.5$$

$$E_{31} = \frac{RT_3 \times CT_1}{GT} = \frac{200 \times 800}{3100} = 51.6$$

$$E_{32} = \frac{RT_3 \times CT_2}{GT} = \frac{200 \times 1000}{3100} = 64.5$$

$$E_{33} = \frac{RT_3 \times CT_3}{GT} = \frac{200 \times 600}{3100} = 38.7$$

$$E_{34} = \frac{RT_3 \times CT_4}{GT} = \frac{200 \times 700}{3100} = 45.2$$

Calculation of Chi-square Value

Cell	Observed Frequency (O)	Expected Frequencies (E)	O – E	$(O - E)^2$	$\frac{(O-E)^2}{E}$
E_{11}	620	593.5	26.5	702.25	1.183
E_{12}	750	741.9	8.1	65.61	0.088
E_{13}	400	445.2	– 45.2	2043.04	4.589
E_{14}	530	519.4	10.6	112.36	0.216
E_{21}	130	154.8	– 24.8	615.04	3.973
E_{22}	200	193.5	6.5	42.25	0.218
E_{23}	140	116.1	23.9	571.21	4.919
E_{24}	130	135.5	– 5.5	30.25	0.223
E_{31}	50	51.6	– 1.6	2.56	0.049
E_{32}	50	64.5	– 14.5	210.25	3.260
E_{33}	60	38.7	21.3	453.69	11.723
E_{34}	40	45.2	– 5.2	27.04	0.598
			0		31.039

$$\therefore \quad \chi^2 = \Sigma\left(\frac{(O-E)^2}{E}\right) = 31.093$$

Degree of freedom,

$$d.f = (3 - 1)(4 - 1)$$
$$= 2 \times 3 = 6$$

At 5% level of significance χ^2 value for d.f = 6 is 12.6. Since $\chi^2_{cal} \leq \chi^2_{table}$ (31.039 > 12.6)

Decision

Null hypothesis is rejected which means quality of product is dependent on machines.

4.7.6 Chi square test for goodness of fit

Q21. Discuss briefly the uses of Chi Square Test as a test of goodness of fit. State the conditions to be satisfied for the applicability of the test.

(OR)

Discuss the Chi Square Test of Goodness of Fit of a theoretical distribution to an observed frequency distribution. State the conditions for the validity of chi square test.

Ans :

χ^2 test help us to find out how well the assumed theoretical distribution fit to the observed data. When some theoretical distribution is fitted to the given data, the statistician or managers will be interested in knowing as to how well this distribution fits with the observed data.

This method of χ^2 test helps in answering this question.

If the calculated value of χ^2 is less than the table value at a certain level of significance, the fit is considered to be good one i.e., divergence between the observed and expected frequencies is attributable to fluctuations of sampling. But if the reverse occurs, the fit is not considered to be a good one. In short,

$\chi^2_{cal} < \chi^2_{table} \Rightarrow$ Good fit

$\chi^2_{cal} < \chi^2_{table} \Rightarrow$ Not a good fit.

If O for i = 1, 2, ...n is a set of observed (experimental) frequencies and E. (i = 1, ...n) is the corresponding set of theoretical frequencies then $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$ with the condition that,

$\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i = N$ = Total frequency follows, χ^2- Distribution with (n – 1) d.o.f.

Steps for Test of Goodness of Fit

1. **Null Hypothesis :** Good fit exists between the theoretical distribution and given data.
2. **Alternative Hypothesis :** No good fit.
3. **Level of Significance** is α.
4. **Critical Region :** Reject null hypothesis if $\chi^2 > \chi^2_\alpha$ with v d.o.f. i.e., theoretical distribution is a poor fit.
5. **Computations :** $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$
6. **Decision :** Accept null hypothesis, if % < Xa. i.e., the theoretical distribution is a good fit to the data.

Conditions for Validity of χ^2 -Test

1. Sample size n should be large (i.e., $n \geq 50$)
2. If individual frequencies O_i. or e_i are small say $O_i < 10$ then combine neighboring frequencies so that combined frequency o. (or e) is ≥ 10.
3. The number of classes k should be neither too small nor too large. In general $4 \leq k \leq 16$.

Degrees of Freedom for X -Distribution

Let k be the number of terms in the formula

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - e_i)^2}{e_i}$$

The d.o.f for X is :

(a) $v = k - 1$ if e_i can be calculated without estimating population parameters from sample statistics.

(b) $v = k - 1 - m$ if e_i can be calculated only by estimating m number of population parameter from sample statistics.

15. Five coins are tossed 3200 times and the number of heads appearing each time is noted. At the end, the following results were obtained.

No. of heads	0	1	2	3	4	5
Frequency	80	570	1100	900	500	50

Test the goodness of fit to determine whether the coins are unbiased. Use 5% LOS.

Sol :

Null Hypothesis : H_0 = The coins are unbiased

Alternative Hypothesis : H_1 = The coins are biased

x	f
0	80
1	570
2	1100
3	900
4	500
5	50
	$\Sigma f = 3,200$

Step-1: Calculate the Values of 'p' and 'q'

In case of tossing a coin,

Probability of getting a head, $p = \frac{1}{2}$

Probability of not getting a head, $q = \frac{1}{2}$

$p(r) = {}^nC_r p^r q^{n-r}$

As there are 6 terms, 'n' is one less than the number of terms.

$\therefore \quad n = 5, N = \Sigma f = 3,200$

Step-2

Compute expected frequencies using binomial probability law,

$$p(r) = {}^nC_r p^r q^{n-r}$$

Where, $r = 0, 1, 2, 3, 4, 5$

$$p = \frac{1}{2}, q = \frac{1}{2}$$

Multiply each of the expected probabilities with the total frequencies (N or Σf) to obtain expected frequency in each case.

$f(r) = N.p(r)$

$N = 3200$

No. of Heads (r)	$p(r) = {}^nC_r, p^r q^{n-r}$	Expected Frequencies $f(r) = N \times p(r)$
f(0)	$^5C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3 = 1\left(\frac{1}{2}\right)^2 = \frac{1}{32}$	$3200 \times \frac{1}{32} = 100$
f(1)	$^5C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 = 5\left(\frac{1}{2}\right)^5 = \frac{1}{32}$	$3200 \times 5 \times \frac{1}{32} = 500$
f(2)	$^5C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = 10\left(\frac{1}{2}\right)^5 = 10 \times \frac{1}{32}$	$3200 \times 10 \times \frac{1}{32} = 1000$
f(3)	$^5C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = 10\left(\frac{1}{2}\right)^5 = 10 \times \frac{1}{32}$	$3200 \times 10 \times \frac{1}{32} = 1000$
f(4)	$^5C_3\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 = 5\left(\frac{1}{2}\right)^5 = 5 \times \frac{1}{32}$	$3200 \times 5 \times \frac{1}{32} = 500$
f(5)	$^5C_3\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 = 1\left(\frac{1}{2}\right)^5 = 1 \times \frac{1}{32}$	$3200 \times \frac{1}{32} = 100$

Step-3: Calculation of Chi-square (x^2)

Number of Heads	O	E	O – E	$\frac{(O-E)^2}{E}$
0	80	100	– 20	$\frac{400}{100} = 4$
1	570	500	70	$\frac{4,900}{500} = 9.8$
2	1100	1000	100	$\frac{10,000}{1,000} = 10$
3	900	1000	– 100	$\frac{10,000}{1,000} = 10$
4	500	500	0	0
5	50	100	– 50	$\frac{2,500}{100} = 25$
				58.8

$$\therefore \quad \chi^2 = \Sigma\left(\frac{O-E}{E}\right) = 58.8$$

Level of significance (a) = 0.05

Degree of freedom = n – l = 6 – l = 5

Table of χ^2 at 5 d.f and 0.05 is 11.07

Decision

Since, the calculated χ^2 i (58.8) is higher than critical value (χ^2 = 11.07), we reject null hypothesis. Therefore, it can be concluded that coins are biased and distribution given is not a good fit to data.

16. The shopkeeper feels that daily demand of a product follows uniform distribution. The observed frequencies of demand values are as follows:

Observed frequencies of daily demand

Demand (Units)	Observed Frequency (O_1)	Demand (Units)	Observed Frequency (O_1)
20	13	26	12
21	10	27	10
22	7	28	14
23	10	29	9
24	6		
25	9		

Check whether the given data follow uniform distribution at a significance level of 0.05.

Sol : **(Imp.)**

Null Hypothesis (H_0): The demand of a product is uniformly distributed.

Alternative Hypothesis (H_a) : The demand of a product is not uniformly distributed.

Level of significance α = 0.05

Total of observed frequencies = 100

Number of demand value = 10

Expected frequency for each demand = $\frac{100}{10}$ = 10

Computation of χ^2 Value TTT . Observed Expected

S.No.	Demand Value	Frequency (O)	Frequency (E)	(O – E)	$(O - E)^2$	$(O - E)^2/E$
1	20	13	10	3	9	0.9
2	21	10	10	0	0	0
3	22	7	10	– 3	9	0.9
4	23	10	10	0	0	0
5	24	6	10	– 4	16	1.6
6	25	9	10	– 1	1	0.1
7	26	12	10	2	4	0.4
8	27	10	10	0	0	0
9	28	14	10	4	16	1.6
10	29	9	10	– 1	1	0.1
	Total	100	100	0		5.6

$$\therefore \quad \chi^2 = \Sigma\left(\frac{(O-E)^2}{E}\right)$$

Degree of freedom = 10 – 1 = 9

Tabulated value of χ^2 for 9 d.f at 5% level of significance is 16.9.

Decision

Since calculated value of χ^2 = 5.6 is less than tabulated value 16.9, null hypothesis is accepted. Hence, it may be concluded that demand of a product is uniformly distributed.

17. The following mistakes per page were seen in a book.

No. of Mistakes Per Page	0	1	2	3	4
No. of Pages	211	90	19	5	0

Fit a poisson distribution and test the goodness of fit.

Sol :

Null Hypothesis (H_0) : Poisson distribution is good fit for data.

Alternative Hypothesis (H_a) : Poisson distribution is not good fit for data.

Compute Mean of Given Distribution

x	f	fx
0	211	0
1	90	90
2	19	38
3	5	15
4	0	0

$N = \Sigma f = 325$

$\Sigma fx = 143$

$$\bar{x} = \frac{\Sigma fx}{N} = \frac{143}{325} = 0.44$$

Fitting a poisson distribution to given data by taking parameter (m) equal to mean of distribution.

$m = \bar{x} = 0.44$

Compute expected frequencies using poisson law $325 \times e^{-0.44} (0.44)^r$

$$F(r) = n.p(r) = \frac{325 \times e^{-0.44}(0.44)^r}{r!}; r = 0, 1, 2, 3, 4$$

$$F(0) = \frac{325 \times e^{-0.44}(0.44)^0}{0!} = 325 \times 0.644 = 209.3$$

$F(1) = m \times F(0) = 0.44 \times 209.3 = 92.1$

$$F(2) = \frac{m}{2} \times F(1) = \frac{0.44}{2} \times 92.1 = 20.3$$

$$F(3) = \frac{m}{3} \times F(2) = \frac{0.44}{3} \times 20.3 = 3$$

$$F(4) = \frac{m}{4} \times F(3) = \frac{0.44}{4} \times 3 = 0.3$$

Testing Goodness of Fit

Observed Frequency (O)	Expected Fequency (E)	(O – E)	$(O - E)^2$	$(O - E)^2/E$
211	209.3	1.7	2.89	0.01381
90	92.1	–2.1	4.41	0.04788
19, 5, 0 } 24	20.3, 3, 0.3 } 23.6	0.4	0.16	0.00678
325	325	0		0.06847

$$\therefore \quad \chi^2 = \Sigma\left(\frac{(O-E)^2}{E}\right) = 0.06847$$

Degree of freedom n = 5 required d.f are V = 5 – 1 – 1– 2 = 1

Note:

One d.f is lost for linear constraint, 1 d.f is lost for parameter and 2 d.f are lost for pooling less than 5 frequencies.

The table value of χ^2 for 1 d.f at 5% level of significance is 3.841.

Decision

Since calculated value of χ^2 (0.06847) is less than table value (3.841), it is not significant.

Hence, it may be concluded that poisson distribution ot given data is a good fit.

Short Questions and Answers

1. Define F-test.

Ans :

Meaning

F-Test or F-distribution is a continuous probability distribution used when two different normal population, are sampled. Consider S_1^2 and S_2^2 as the sample variances of different random sample of sizes n_1 and n_2 respectively. These samples are drawn from two different normal population $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, where (μ_1, σ_1^2) and (μ_2, σ_2^2) denotes the mean and variances of Sf and S\ respectively.

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

Inorder to determine whether the samples (S_1^2, S_2^2) are drawn from two different population, having equal variances. It is necessary to compute the ratio of variances related to two independent random sample.

2. Properties of F-distribution.

Ans :

i) The F-distribution curve lies in only first quadrant (Q_1) and is unimodal.

ii) The F-distribution is independent (free) of population parameter and depends only on the degree of freedom (i.e., V_1 and V_2) according to its order.

iii) The F-distribution mode is less than unity (i.e., mode < 1).

3. Association of Attributes.

Ans :

Meaning

Association in statistical language is quiet different from the common meaning of association. Commonly, if two attributes A and B appear together number of times then they can be said to be as associated.

Definition

According to Yule and Kendall, "In Statistics A and B are associated only if they appear together in a greater number of cases than is to be expected, if they are independent."

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.

5. Scope of correlation analysis.

Ans :

(i) One of the variable may be affecting the other

A correlation coefficient calculated from the data on quantity demanded and corresponding price of tea would only reveal that the degree of association between them is very high. It will not give us any idea about whether price is affecting demand of tea or vice-versa. In order to know this, we need to have some additional information apart from the study of correlation.

(ii) The two variables may act upon each other

Cause and effect relation exists in this case also but it may be very difficult to find out which of the two variables is independent.

(iii) The two variables may be acted upon by the outside influences:

In this case we might get a high value of correlation between the two variables, however, apparently no cause and effect type relation seems to exist between them.

6. What are the Properties of Correlation?

Ans :

Properties

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.

7. Proportion method.

Sol :

If there is no relationship of any kind between two attributes A and S-we expect to find the same proportion of A's amongst the B's as amongst the P's. Thus, if a coin is tossed we expect the same proportion of heads irrespective of whether the coin is tossed by the right hand or the left hand.

Symbolically, two attributes may be termed :

(i) Independent if $\frac{(AB)}{B} = \frac{(A\beta)}{(\beta)}$

(ii) Positively associated if $\frac{(AB)}{B} > \frac{(A\beta)}{(\beta)}$

(iii) Negatively associated if $\frac{(AB)}{B} < \frac{(A\beta)}{(\beta)}$

If the relation (i) holds good the corresponding relations

$$\frac{(\alpha B)}{(B)} = \frac{(\alpha\beta)}{(\beta)}, \frac{(AB)}{A} = \frac{(\alpha B)}{(\alpha)}, \frac{(AB)}{(A)} = \frac{(\alpha\beta)}{(\alpha)}$$

must also hold true.

8. Yule's coefficient of association.

Ans :

The most popular method of studying association is the Yule's Coefficient because here not only we can determine the nature of association, *i.e.,* whether the attributes are positively associated, negatively associated or independent, but also the degree or extent to which the two attributes are associated. The Yule's Coefficient is denoted by the symbol Q and is obtained by applying the following formula:

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

The value of this coefficient lies between ± 1. When the value of Q is + 1 there is perfect positive association between the attributes. When Q is – 1 there is perfect negative association (or perfect dissociation) between the attributes and when the value Q is zero the two attributes are independent.

The coefficient of association can be used to compare the intensity of association between two attributes with the intensity of association between two other attributes.

9. Applications of Chi-square.

Ans :

Chi-square distribution has a number of applications,

i) Chi-square test of goodness of fit.

ii) Chi-square test for independence of attributes.

iii) To test if population has a specified value of variance.

iv) To test equality of several population proportions.

10. Yates's correction.

Ans :

The Yates' corrections, also called Yates' corrections for continuity, are introduced because theoretical chi-square distribution is continuous whereas the tabulated values are based on the distribution of discrete χ^2 statistic. The corrections has the effect of reducing the calculated value of χ^2 as continuous compared to the corresponding value without correction.

In a special case of 2 x 2 contingency table the approximation may be improved, and bias arising out of the use of small theoretical frequencies may be reduced, by means of a correction proposed by F. Yates in 1934. The correction involves the reduction of the deviation of observed from theoretical frequencies which of course reduces the value of χ^2. The working rule for the application of the correction is : adjust the observed frequency in each cell

of the 2×2 table in such a way as to reduce the absolute deviation of the $\frac{1}{2}$ observed from the theoretical frequency for that cell by adjustments for all the cells are to be made without changing the marginal totals. This operation will increase f_0, that is observed frequency, by $\frac{1}{2}$ in each of two cells, and will reduce f_0 by $\frac{1}{2}$ in each of two cells.

Exercise Problems

1. Four experimenters determine the moisture content of samples of a powder, each man taking a random sample from each of the six consignments. These assessments are given in Table:

Observer	Consignment					
	1	2	3	4	5	6
1	9	10	9	10	11	11
2	12	11	9	11	10	10
3	11	10	10	12	11	10
4	12	13	11	14	12	10

Perform an analysis of variance on these data and discuss whether there is any significant difference between consignments or between observers.

If there is significant difference between the levels of any of the factor (s) of variation, then determine the corresponding pairs of sample means which differ significantly. Use a = 0.05. [Refer Similar, Q26]

[Ans : F_A = 5.03, F_B = 2.23, Reject Null Hypothesis (H_{0A}) and Accept Null Hypothesis (H_{0B})]

---

2. A random sample of size 20 yields a mean of 40 and variance of 25. Test the hypothesis that the population S.D. is 8.

[Given χ^2 0.05 = 28.87 for 18 d.f. ; = 30.14 for 19 d.f; = 31.41 for20d.f.]

[Ans : H_0: σ = 8 ; χ^2 = 7.81, Not Significant]

---

3. Suppose that, in a public opinion survey answers to the question

(a) Do you drink?

(b) Are you in favour of local option on sale of liquor?

Were as given in the table

Question (b)	Question (a)		Total
	Yes	No	
Yes	56	31	87
No	18	6	24
Total	74	37	111

Can you infer that opinion on local option is dependent on whether or not an individual drinks?

[Ans : χ^2 = 0.9568 (Not significant) null hypothesis is accepted]

4. The adjoining data represent the number of units of production per day turned out by 5 different workmen using different types of machines.

Workmen	Machine Type			
	A	B	C	D
1	46	40	49	38
2	48	42	54	45
3	36	38	46	34
4	35	40	48	35
5	40	44	51	41

(i) Test whether the mean productivity is the same for the four brands of machine types.

(ii) Test whether 5 different workmen differ with respect to productivity.

[Ans : F (Workmen) = 8.20, Significant; F (Machines) = 19.20, Significant]

Choose the Correct Answers

1. When observed and expected frequencies completely coincide, χ^2 will be _____. [d]
   (a) +1 (b) –1
   (c) Greater than 1 (d) Zero

2. For v = 2, $\chi^2_{0.05}$ equals [b]
   (a) 5.90 (b) 5.99
   (c) 5.55 (d) 5.95

3. The calculated value of χ^2 is _____. [a]
   (a) Always positive (b) Always negative
   (c) Can be either positive or negative (d) None

4. For degrees of freedom greater than 30 _____. [c]
   (a) $2\chi^2 - \sqrt{2\nu} - 1$ (b) $\sqrt{3\chi^2} - 2\nu - 1$
   (c) $\sqrt{2\chi^2} - \sqrt{2\nu} - 1$ (d) $\sqrt{2\chi^2} - 2\nu - 1$

5. Coefficient of contingency is given by _____. [c]
   (a) $\chi^2 / N + \chi^2$ (b) $\sqrt{\chi^2} / N + \chi^2$
   (c) $\sqrt{\chi^2} / N + \chi^3$ (d) $\chi^2 / N^2 + \chi^2$

6. The Yate's corrections are generally made when number of degrees of freedom is _____. [d]
   (a) 5 (b) Greater than 5
   (c) 4 (d) 1

7. In a contingency table, degrees of freedom are determined by _____. [a]
   (a) (r – 1); (c – 1); (b) (r – 1); (c + 1);
   (c) (c – 1); (r) (d) (cr);

8. The number of degrees of freedom in a 3 × 3 contingency table is _____. [b]
   (a) 8 (b) 4
   (c) 3 (d) 1

9. The χ^2 test was devised as : _____. [d]
   (a) Fisher (b) Gauss
   (c) Laplace (d) Karl Pearson

10. The χ^2 test is defined as : _____. [c]
   (a) $\chi^2 = \Sigma\frac{(O-E)}{E}$ (b) $\chi^2 = \Sigma\frac{(O+E)}{E}$
   (c) $\chi^2 = \Sigma\frac{(O+E)^2}{E}$ (d) $\chi^2 = \Sigma\frac{(O+E)^2}{E}$

Fill in the Blanks

1. _____ is a continuous probability distribution used when two different normal population, are sampled.
2. ANOVA is the acronym for _____.
3. F* test as it was developed by _____ in 1920's.
4. _____ is used in education, industry, business, psychology fields mainly in their experiment design.
5. _____ is the study of the linear relationship between two variables.
6. The most popular method of studying association is the _____ Coefficient
7. The square of a standard normal variable is called a _____.
8. _____ describes the magnitude of discrepancy between theory and observations.
9. _____ is very useful to economists to study the relationship between variables.
10. The value of correlation 'r' varies between _____.

ANSWERS

1. F-distribution
2. Analysis of Variance
3. R.A Fisher
4. Anova
5. Correlation
6. Yule's
7. Chi-square
8. χ^2
9. Correlation
10. [–1, +1]

One Mark Answers

1. One Way ANOVA.

Ans :

In one way classification/single factor ANOVA, only one factor is considered and its effect on elementary units is observed, i.e., data are classified according to only one criterion. For instance, Yield of crop affected by type of seed only.

2. 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. For instance, Yield of crop affected by type of seed as well as type of fertilizer.

3. Contingency table.

Ans :

Qualitative data are often classified into more than two classes, *i.e.*, attribute A may be classified not as 'A' and 'not A', but as A_1, A_2, A_3, etc. Similarly, another attribute B may be subdivided into B_1, B_2, B_3, etc.

4. Assumptions of chi-square.

Ans :

i) Sample observation data must be independent of each other.

ii) Random sampling from specified population to give sample data.

iii) Data should not be in percentage or ratio form but in original units to make comparison easy.

iv) Sample size should have atleast 50 observations.

5. Limitations of Chi Square Test.

Ans :

i) Frequencies of non-occurrence should not be omitted for binomial or multinomial events.

ii) The formula presented for χ^2 statistics in terms of frequencies. Hence and attempt should not be made to compute on the basis of proportions or another derived measures.

UNIT V

Statistical Decision Theory, Game Theory and Linear Programming:

Statistical Decision Theory: Nature of Decision - State of Nature – Pay off Tables - Expected Pay off - Expected Opportunity Loss – Value of Perfect Information – Types of Decision Situation – Choice of Decision Criteria – Decision Tree Analysis – Decision Making under Uncertainty (including simple problems).;

Game Theory: Characteristics of Game Theory – Two Persons Zero Sum Game - Maximum and Minimax Strategies – Saddle Point – Dominating Strategy – Mixed Strategy - Limitations of Game Theory (including simple problems with Analytical Formulae and Graphical Methods).

Linear Programming: Linear Programming: Meaning - Requirements for application - Assumptions - Advantages - Application of LP - Formulation of LP problems (including simple problems). Graphical Solutions of LP problems with two variables only (including simple problems).

5.1 Statistical Decision Theory

Q1. Explain the concept of Statistical Decision Theory.

Ans : **(Imp.)**

Meaning

The businessman has to operate in an atmosphere of uncertainty and has to select the best course out of several alternative courses of action that may be available to him. In earlier days, decisions were made mainly on personal judgement. However, these day& judgement is combined with several quantitative techniques and the best action is arrived at in a given situation.

The hypothesis testing is sometimes called classical decision theory. In hypothesis testing, the statistical decision is to either accept or reject the null hypothesis.

Procedure

The classical decision procedure has the following three major defects:

(i) **Firstly,** it provides for only two possible actions which correspond to either acceptance or rejection of the null hypothesis. Also, it allows only two states of the parameter being tested, that is, those parameter values that make the null hypothesis true and those values for which the null hypothesis is false.

(ii) **Secondly,** classical decision procedure does not recognize the validity of an information pertaining to the decision that does not exist in the form of empirical data that result from a process of sampling.

(iii) **Thirdly,** there are real economic consequences that result from making a wrong decision. Although such consequences might be considered by the decision-maker using the classical procedure in selecting the significance level for the test, these considerations are never an explicit part of the decision model or procedure.

5.1.1 Nature of Decision

Q2. Explain the nature of decision theory.

Ans :

Decision theory which is defined as the process of logical and quantitative analysis of all factors that influences the decision problem, assists the decision-maker in analyzing these problems with several courses of action and consequences.

The basic objective of decision theory is to provide a method of natural decision-making wherein data concerning the occurrence of different outcomes (consequences) may be evaluated to enable the decision-maker to identify the best course of action.

Decision models are classified according to degree of certainty. The scale of certainty can range from complete certainty to complete uncertainty. The region which falls between these two extreme points corresponds to the decision making under risk probabilistic model) as shown in figure below.

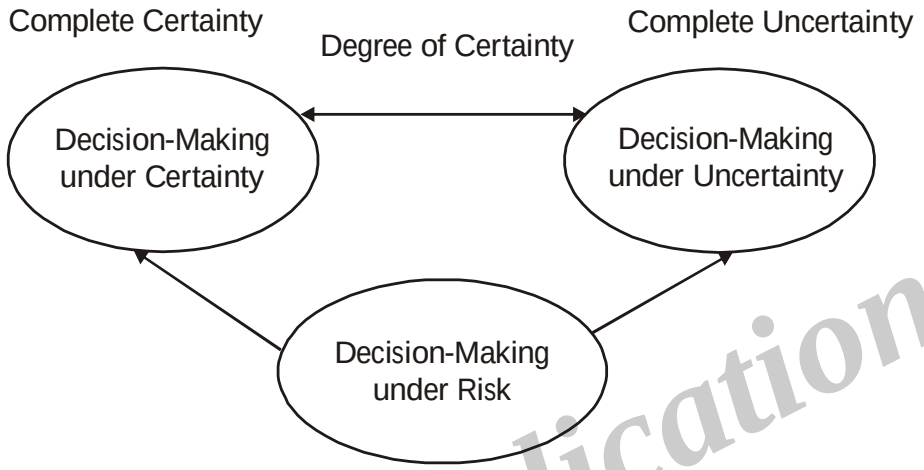


Fig.: Decision Making Situations

The logical and quantitative method of analyzing all the factors affecting the decision problem in assisting the decision maker to analyze these problems with several procedures of action and results is the definition of decision theory.

To identify the best possible action by the decision maker and enable him to evaluate the available data regarding the happening of the various events, to provide him with a method of natural decision-making is the foremost objective of the decision theory. The classification of decision models is in accordance with the degree of certainty. The scale of certainty varies from complete certainty to complete uncertainty. The specific region falling between the two extremities corresponds to the decision making under the possibility of risk.

Q3. Explain the steps involved in decision making.

(OR)

Elaborate the various steps involved in decision making.

Ans : **(Imp.)**

Step1. Define the Problem

The first action has to be taken by the decision maker is to define the problem clearly. In decision making, it is a very crucial step as ambiguously defined problems will only produce poor results. It is well said "a problem well defined is half solved". Thus, problems which are not clearly defined, will hinder the decision maker to decide correctly.

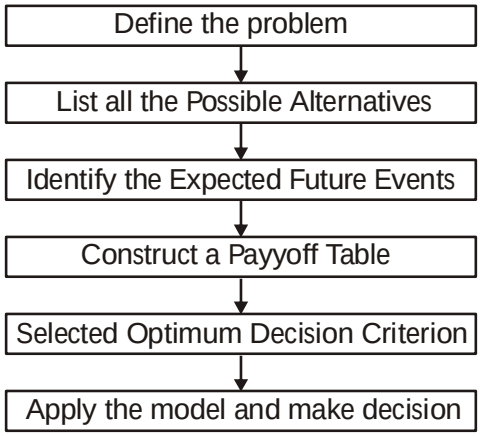


Fig.: Steps of Decision Making

Step2. List all the possible alternatives

For a problem that has been clearly defined its second step will be to make a list of those possible options that have been considered for its decision. For example, Following are the three options for the company :

(i) Expansion of the existing plant.

(ii) Construction of a new plant.

(iii) Hiring production for further demand.

Step3. Identify the Expected Future Events

Listing of all the possible occurrence of future events is the third step for the decision maker. It is very much possible to identify the happening of the events, but the difficulty is in the identification of the occurrence of any particular event. In decision theory, those events which are not under the control of decision maker are known as states of nature. Definitely, only one event will occur from the given list. For example, a manufacturing company having greatest uncertainty about the demand for the product. The future events related to the demand may be :

(i) High demand.

(ii) Moderate demand.

(iii) Low demand.

(iv) No demand.

Step4. Construct a Payoff Table

The decision maker constructs a payoff table for each possible combination of alternative course of action and state of nature (table comprising of profit, benefit, etc.)

Step5. Select Optimum Decision Criterion

After the construction of the payoff table, the decision maker has to decide the choice of the best criterion, which yields largest payoff. It may be economic, quantitative or qualitative.

For example, market share, profit, fragrance of a perfume, etc.

Step6. Apply the Model and Make Decision

The selected model is finally applied by the decision maker, which facilitates in decision-making.

5.2 STATE OF NATURE – PAY OFF TABLES - EXPECTED PAY OFF

Q4. Explain the various elements of decision making.

Ans : **(Imp.)**

1. Decision-Maker

The decision maker is either an individual or a group of individuals, who are responsible to choose a suitable course of action from the available courses of action.

2. Goals

Certain objectives that a decision maker is willing to attain through his actions are known as goals.

3. Preference on Value System

This refers to the norms in deciding the best course of action by the decision maker that includes maximization of income, utility, profit, etc.

4. Acts/Courses of Action

The courses of action or decision choices is also known as acts. Hence, for any difficult situation every possible plans of action must be included. Whether these possible courses of action are large or small, it should be emphasized that these must be under the control of the decision maker.

5. States of Nature

The decision-maker must develop a complete list of possible future events before the application of decision theory. Though, there is no direct control of the decision-maker over the happening of a particular event. These future events are known as states of nature and it is assumed to be mutually exclusive and collectively exhaustive.

These states of nature can be numerically descriptive, as demand for some units of a particular item or a non- numerical description of an employees' strike.

6. Payoff

Payoffs are also known as conditional values or profits, and its effectiveness is linked with a particular combination of a plan of action and state of nature.

7. Payoff Table

A payoff table is a record of the states of nature (outcomes), that is mutually exclusive and collectively exhaustive with a set of likely courses of action. The payoff is calculated for each combination of nature and course of action.

The association of the weighted profit with the specified combination of state of nature and course of action is acquired by the multiplication of the payoff, for that specific state of nature and course of action with the probability of happening of the given outcome. The table shown below is an example of payoff table. According to this table,

States of Nature	Courses of Action			
	S_1	S_2	S_1	S_n
O_1	a_{11}	a_{12}...	a_{ij}...	...a_{1n}
O_2	a_{21}	a_{22}...	...a_{ij}...	...a_{2n}
	⋮	⋮	⋮	⋮
O_i	A_{i1}	A_{i2}...	...a_{ij}...	...a_{1n}
	⋮	⋮	⋮	⋮
O_m	a_{m1}	A_{m2}...	a_{mi}...	...a_{mn}

8. Opportunity Loss

The failure of not choosing the most positive course of action or approach is responsible for incurring an opportunity loss. For each state of nature or outcome, opportunity loss values are calculated individually that is, by initially finding the most positive course of action for that particular outcome. Then by stating the difference between that payoff value for that course of action and the payoff value for the best feasible course of action should be selected.

	A_1 : 18	A_2 : 19	A_3 : 20	A_4 : 21	A_5 : 22	A_6 : 23
D_1: 18	0	50	100	150	200	250
D_2: 19	20	0	50	100	150	200
D_3: 20	40	20	0	50	100	150
D_4: 21	60	40	20	0	50	100
D_5: 22	80	60	40	20	0	50
D_6: 23	100	80	60	40	20	0

Table. : Regret

5.3 EXPECTED OPPORTUNITY LOSS

Q5. Explain briefly about Expected Opportunity Loss (EOL).

(OR)

Define EOL. Explain the various steps for calculating EOL.

Ans : **(Imp.)**

Meaning

An alternative approach that refers to maximization of Expected Monetary Value (EMV) and minimization of Expected Opportunity Loss (EOL) is also known as expected value of regret. The difference between the highest profit for a state of nature and the actual profit which is obtained for the specific course of action is defined as EOL.

Hence, the amount of payoff that is lost due to the rejection of a course of action, which is having the greatest payoff for the state of nature that has actually appeared, is referred as EOL. That course of action is recommended for which EOL is minimum.

Results obtained by EMV criterion and by EOL, which is an alternative decision criterion for making decision under risk area will always be the same. Hence, only one of the two methods should be applied for reaching a decision. The mathematical description is as follows:

$$EOL(N_1) = \sum_{i=1}^{m} 1_{ij} P_i$$

Where,

1_{ij} = opportunity loss due to state of nature, N_i and course of action, S_j

P_i = probability of occurrence of state of nature, N_i

Steps

Step1:

Preparation of a conditional profit table depicting that course of action, state of nature as well as associated probabilities.

Step 2:

Calculate Conditional Opportunity Loss (COL) value for each state of nature and subtract by each payoff from the maximum payoff of each event.

Step 3:

Calculation of EOL for each course of action that is multiplied by the probability of each state of nature with the value of COL and then adding it up.

Step 4:

That course of action is selected for which the EOL value is minimum.

PROBLEMS

1. A conditional profit matrix is given below,

States of Nature	Probabilities	Course of Action		Condition Profit	
		A_1	A_2	A_3	A_4
N_1	0.15	200	160	130	100
N_2	0.20	200	210	180	150
N_3	0.30	200	210	230	190
N_4	0.35	200	210	230	250

From this given matrix calculate conditional opportunity loss values and EOL.

Sol :

Calculating Conditional Opportunity Loss (COL) and Expected Opportunity Loss (EOL) values.

States of nature	Probabilities (1)	Course of action Conditional opportunity loss (COL)				Courses of action Expected opportunity loss (EOL)			
		A_1 (2)	A_2 (3)	A_3 (4)	A_4 (5)	A_1 (1)×(2)	A_2 (1)×(3)	A_3 (1)×(4)	A_4 (1)×(5)
N_1	0.15	0 [200–200]	40 [200–160]	70 [200–130]	100 [200–100]	0	6	10.5	15
N_2	0.20	10 [210–200]	0 [210–210]	30 [210–180]	60 [210–150]	2	0	6	12
N_3	0.30	30 [230–200]	20 [230–210]	0 [230–230]	40 [230–190	9	6	0	12
N_4	0.35	50 [250–200]	40 [250–210]	20 [250–230]	0 [250–250]	17.5	14	7	0
		Expected opportunity Loss (EOL)				28.5	26	23.5	39

The minimum expected opportunity loss is 23.5 which corresponds to course of action A_3. The optimum course of action is A_3 as per EOL criterion.

5.3.1 Value of Perfect Information

Q6. Briefly explain about the determination of EVPI and factors involved in the design of marketing research experiments.

Ans : **(Imp.)**

Meaning

Valuing information is very important especially when the information is purchased from a third party. Decision makers compute the expected profit with perfect information and compare it with the cost of acquiring

information to see if the information is worth buying or not. Hence, expected value of perfect information is the maximum cost a decision maker is agreeable to pay to acquire the perfect information about the happening or non-happening of an event.

Thus, Expected Value of Perfect Information = Expected Profit with Perfect Information -Expected Monetary Value

Factors

Before designing the marketing research experiment, an appropriate and decision-oriented research design must be selected. The various factors to be taken into consideration while selecting an experimental design are,

(a) Cost of the design

(b) Time consumed in the research design

(c) Secrecy required.

(a) Cost of the Design

A decision maker must choose the experimental design which involves less cost. If the cost involved in using an experimental design is higher, then the loss associated with such experimental design will be higher. So, the best decision would be selecting the experimental design with minimum cost.

(b) Time Consumed in the Research Design

Today's competitive scenario requires management to take quick decisions. Lengthy experimental designs consuming long time should be avoided.

(c) Secrecy Required

The selection of experimental design also get influenced by the factor 'secrecy'. If the decision makers want to hide themselves from the competitors then selecting the simulation method or laboratory experiments would be a good choice.

Q7. Explain the posterior analysis of decision-making. Define the expected value of sample information. Give a measure of efficiency of EVSI.

Ans :

Meaning

The probabilities of an event before the collection of new information is known as 'Priori probabilities'.

Posterior probabilities are the revised priori probabilities that have been derived after using the new information. It is also known as 'Inverse' or "Revised' probabilities.

If, A_1, A_2, A_3... A_n are mutually exclusive and collectively exhaustive events, then

$P(A_1)$, $P(A_2)$, $P(A_3)$,...,$P(A_n)$ are the priori probability,

'B' is an event such that $P(B) \neq 0$ whose conditional probabilities are represented as,

$P(B/A_1)$, $P(B/A_2)$, $P(B/A_3)$,.....,$P(B/A_n)$

It should be noted that the conditional probabilities are known with the help of this data we have to calculate posterior probabilities.

$\therefore$ Posterior probabilities can be calculated by using the following formula,

$$P(A_i / B) = \frac{P(A_i \cap B)}{\sum_{i=1}^{n}(A_i \cap B)} = \frac{P(B / A_i)P(A_i)}{\sum_{i=1}^{n}P(B / A_i)P(A_i)}$$

(or)

$$P(A_i / B) = \frac{P(A_i \cap B)}{P(B)}$$

Where,

$P(A_i \cap B)$ is the joint probability of A. and B events

A_i and B events

$P(A_i)$ is the priori probability and

$P(B/A_i)$ is the conditional probability.

$$\therefore P(B) = \sum_{i=1}^{n}P(B / A_i)P(A_i)$$

Prior probabilities of high and low demand are revised in order to determine posterior proababilities in first position. Hence, a decision maker uses this posterior probabilities to make a decision. This analysis is known as posterior analysis.

Expected Value of Sample Information (EVSI)

A sample information is a information which is revised on the basis of prior probabilities, Hence, determination of Expected Value of Sample Information (EVSI) is required. EVSI is calculated in the following manner,

Expected Value of Sample Information (EVSI) = Expected pay-off with Sample information - Expected pay-off without sample information.

Efficiency of EVSI

While developing a new product, a decision maker have many kinds of sample information through survey focus group or other market research techniques. However, these information may hot be perfect all the time. In such situation, sample information can be evaluated by relating EVSI with EVP1. When information sample is perfect, the efficiency will be 100%. The efficiency index (EI) of sample information can be calculated as,

$$\text{Efficiency of Sample Information (EI)} = \frac{EVSI}{EVPI} \times 100$$

Q8. Briefly explain the difference between expected opportunity loss and expected value of perfect information.

Ans :

The following are the differences between EOL and EVPI,

S.No	Nature	Expected Opportunity Loss (EOL)	Expected Value of Perfect Information (EVPI)
1.	Meaning	The expected opportunity loss (EOL) refers to the expected difference between the conditional values or profits of the best course of action and the course of action taken.	The expected value of perfect information is the maximum cost a decision maker is agreeable to pay to acquire the perfect information about the happening or non-happening of an event.
2.	Alternative	EOL is an alternative to EMV (Expected Monetary Value).	EVPI is not an alternative to EMV.
3.	Formulae	EOL is symbolically expressed as, $EOL(\text{State of nature, } N_j) = \sum_{i=0}^{n} P(O)_i L_{ij}$	The formula for calculating EVPI can be expressed as, EVPI = Expected profit with perfect information expected monetary value.

PROBLEMS

2. Daily demand for loaves of bread at a grocery store are given by the following probability distribution.

X	100	150	200	250	300
p(x)	0.20	0.25	0.30	0.15	0.10

If a loaf is not sold the same day, it can be disposed of at ` 1 at the end of the day, otherwise the price of a fresh loaf is ` 11. The cost per loaf to the store is ` 2.50. Assuming stock level is restricted to one of the demand levels, how many loaves should be stocked daily?

Sol :

Marginal Profit (MP) = Selling Price – Cost

= 11–2.50

= 8.50

Marginal Loss (MC) = Cost Price – Cost of Unsold

= 2.50 – 1

= 1.50

Conditional Profit (Payoff) = (Marginal Profit × Loaves Sold) – (Marginal Loss × Loaves Not Sold)

= (11– 2.50) (Loaves Sold) – (2.50 – 1) (Loaves Unsold)

$$\begin{cases} 8.50\ D & \text{.... if } D \geq N \\ (11-2.50)0 - 1.50(N-D) = 100 - 1.5N & \text{.... if } D < N \end{cases}$$

In the table below, the resulting conditional profit values and corresponding expected payoffs are calculated.

Calculation of Conditional Profit Table and Expected Payoffs

state of Nature	Probability p(x) (1)	Conditional Profit (`) Due to Course of Action					Expected Payoff (`) Due to Course of Action				
		100 (2)	150 (3)	200 (4)	250 (5)	300 (6)	100 (1)×(2)	150 (1)×(3)	200 (1)×(4)	250 (1)×(5)	300 (1)×(6)
100	0.20	850	775	700	625	550	170	155	140	125	110
150	0.25	850	1275	1200	1125	1050	212.5	318.75	300	281.25	262.5
200	0.30	850	1275	1700	1625	1550	255	382.5	510	487.5	465
250	0.15	850	1275	1700	2125	2050	127.5	191.25	255	318.75	307.5
300	0.10	850	1275	1700	2125	2550	85	127.5	170	212.5	255
Expected Monetary Value (EMV)							**850**	**1175**	**1375**	**1425**	**1400**

As the maximum EMV of `1425 corresponds to the course of action 250, the retailer should stock 250 loaves every day.

3. Ramana often flies from Chennai to Hyderabad. He can use the airport bus which cost ` 250 but if he takes it, there is a 0.08 chance he will miss the flight. The stay in a hotel costs ` 2,700 with a 0.96 chance of being on time for the flight. For ` 3,500 he can use a taxi which will make 99 percent chance of being on time for the flight. If Ramana catches the plane on time he will conclude a business transaction which will produce a profit of ` 100,000. Otherwise he will lose it. Which mode of transport should Ramana use? Answer on the basic of EMV criterion.

Sol :

Calculation of EMV in Different Situations

Option - 1 Ramana uses Airport bus :

	Cost	Prob.	Expected Value
Catches the flight	1,00,000 – 250 = 250 = 99,750	× 0.92	91,770
Misses the flight	– 250	× 0.08	– 20
Expected Monetary Value			**91,750**

Option-2 Ramana stays in hotel :

	Cost	Prob.	Expected Value
Catches the flight	1,00,000 – 2,700 = 97,300	× 0.96	93,408
Misses the flight	– 2,700	× 0.04	–108
Expected Monetary Value			**93,300**

Option-3 Ramana stays in hotel :

	Cost	Prob.	Expected Value
Catches the flight	1,00,000 – 3,500 = 96,500	× 0.99	95,535
	– 3,500	× 0.01	35
Expected Monetary Value			**95,500**

After considering all modes of transport we can conclude that Ramana should use taxi because it has the highest EMV.

4. An investment of $10,000 in a high risk venture has a 50 - 50 chance over next year of increasing to $14,000 or decreasing to 8,000. Thus the next return can be either $4,000 or - $2,000.

(a) Assuming a risk-neutral investor and a utility scale from 0 to 100, determine the utility of $0 net return on investment and associated indifference probability.

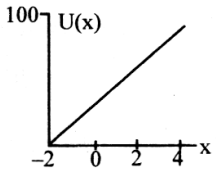
(b) Suppose the two investors A and B have exhibited the following indifference probabilities:

Net Returns (S)	Indifference Probability	
	Investor A	Investor B
– 2000	1.00	1.00
–1000	0.30	0.90
0	0.20	0.80
1000	0.15	0.70
2000	0.10	0.50
3000	0.05	0.40
4000	0.00	0.00

Graph the utility functions for investors A and B, and categorize each investor as either a risk averse person or a risk seeker.

Sol :

(a) Let us establish a utility scale U from 0 to 100 that specifies U (–$2,000) = 0 and U ($ 4,000)



$$\frac{U(0)}{U(4)} = \frac{0-(-2)}{4-(-2)}$$

$$= \frac{2}{6} = \frac{1}{3}$$

$$U(0) = \frac{1}{3}(100) = 33.33$$

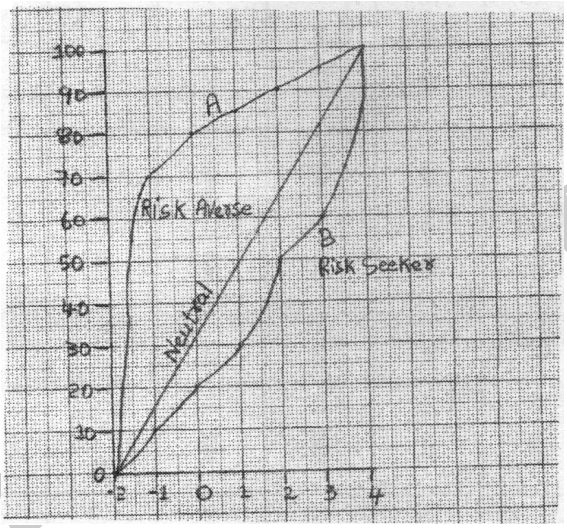
Now, U (0) = pU (–2) + (1 – p) U (4)

= 100 (1 – p)

Thus, for U (0) = 33.33, p = 0.6667 or 66.67

(b)

X	Investor A $U(x)_A$	Investor B $U(x)_B$
– 2	100 (1 – 1) = 0	100 (1 – 1) = 0
– 1	100 (1 – 0.30) = 70	100 (1 – 0.90) = 10
0	100 (1 – 0.20) = 80	100 (1 – 0.80) = 20
1	100 (1 – 0.15) = 85	100 (1 – 0.70) = 30
2	100 (1 – 0.10) = 90	100 (1 – 0.50) = 50
3	100 (1 – 0.05) = 95	100 (1 – 0.40) = 60
4	100 (1 – 0.00) = 100	100 (1 – 0.00) = 100



5. A milk producing co-operative union desires to determine how many kilograms butter it should produce on daily basis to meet the demand. Past records have shown the following pattern of demand.

Quantity demanded (kg.)	15	20	25	30	35	40	45
Days when given demand occurred	4	16	20	80	40	30	10

Assume that the stock levels are restricted to the range 15-45 kg (a multiple of 5 kg) and that butter left unsold at the end of the day must be disposed of due to inadequate storing facilities. Butter costs ` 14 per kg and is sold at ` 20 per kg.

(a) Construct a conditional profit table.

(b) Determine the action alternative associated with the maximization of expected profit.

(c) Determine EVPI.

Sol :

Marginal/Market Profit = 20 – 14 = 6

Marginal/Market Loss = 14

(i) Calculation of Conditional Profit

Conditional Profit = (Market Profit x Butter sold) – (Market Loss × Butter not sold)

= (20 – 14 × Butter sold) – (14 × Butter not sold)

$$= \begin{cases} 6D & ; \ D \geq S \\ 6D - 14(S - D) & ; \ D \leq S \end{cases}$$

Where,

D = Number of units demanded

S = Number of units Stocked

(ii) Computation of Conditional Profit Table and Expected Maximum Profit (EMP)

The dairy firm can produce 30 kg butter and can expect an daily maximum profit of ` 148.

Quantity Demanded (Kg)	Probability	Course of Stock Action Alternative (Kg)												
		15	20	25	30	35	40	45	15	20	25	30	35	40	45	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	[1 × 2]	[1 × 3]	[1 × 4]	[1 × 5]	[1 × 6]	[1 × 7]	[1 × 8]	
15	0.02	90 [15 × 6]	20 [(20 × 15) – (14 × 20)] [300 – 280]	– 50 [300 – (14 × 25)]	– 120 [300 – (14 × 30)]	– 190 [300 – (14 × 35)]	– 260 [300 – (14 × 40)]	– 330 [300 – (14 × 45)]	1.8	0.4	– 1	– 2.4	– 3.8	– 5.2	– 6.6	3
20	0.08	90	120 [(20 × 20) – (14 × 20)] [400 – 280]	50 [400 – (14 × 25)]	– 20 [400 – (14 × 30)]	– 90 [400 – (14 × 35)]	– 160 [400 – (14 × 40)]	– 230 [400 – (14 × 45)]	7.2	9.6	4	– 1.6	– 7.2	– 12.8	– 18.4	68
25	0.10	90	120	150 [(20 × 25) – (14 × 25)] [500 – 350]	80 [500 – (14 × 30)]	10 [500 – (14 × 35)]	– 60 [500 – (14 × 40)]	– 130 [500 – (14 × 45)]	9	12	15	8	1	– 6	– 13	118
30	0.40	90	120	150	180 [(20 × 30) – (14 × 30)] [600 – 420]	110 [600 – (14 × 35)]	40 [600 – (14 × 40)]	– 30 [600 – (14 × 45)]	36	48	60	72	44	16	– 12	148
35	0.20	90	120	150	180	210 [(20 × 35) – (14 × 35)] [700 – 490]	140 [700 – (14 × 40)]	70 [700 – (14 × 45)]	18	24	30	36	42	28	14	138
40	0.15	90	120	150	180	210	240 [(20 × 40) – (14 × 40)] [800 – 560]	170 [800 – (14 × 45)]	13.5	18	22.5	27	31.5	36	25.5	118
45	0.05	90	120	150	180	210	240	270 [(20 × 45) – (14 × 45)] [900 – 630]	4.5	6	7.5	9	10.5	12	13.5	90
																EMV

(iii) Computation of EVPI

Quantity Demand (kg)	Probability (1)	Conditional Profit (`) (2)	Expected Profit with Perfect Information (`) (1 × 2)
15	0.02	90	1.8
20	0.08	120	9.6
25	0.10	150	15
30	0.40	180	72
35	0.20	210	42
40	0.15	240	36
45	0.05	270	13.5
		EPPI	**189.9**

EVPI = EPPI – EMV

= 189.9 – 148

EVPI = 41.9.

5.4 TYPES OF DECISION SITUATION

5.4.1 Choice of Decision Criteria – Decision Tree Analysis – Decision Making under Uncertainty

Q9. Discuss the concept of choice of decision criteria.

Ans : **(Imp.)**

The nature of the decision criteria would depend upon the type of the decision situation as follows:

1. Under Conditions of Certainty

Under this condition; there is one pay-off for each strategy. The pay-off represents the degree of achievements of the objective, hence the largest pay-off is chosen and the corresponding strategy is selected.

2. Under Conditions of Risk

Under condition of risk, there would be more than one state of nature but the probabilities of their occurrence are known on the basis of past experience. The strategy which gives the maximum pay-off is selected.

3. Under Conditions of Uncertainty

Under conditions of uncertainty, we do not have a set of probabilities for the state of nature. Therefore, for each alternative only pay-offs or utilities are known. But nothing is known about the likelihood of each state of nature. The problem becomes more complex and the personality of the decision maker plays an important role in the selection of the strategy.

The following criteria are generally followed:

(i) Maximin Criterion

The strategy which gives the highest minimum pay-off will be chosen. The basic rationale behind this criterion is that pessimism is not irrational under the state of uncertainty. The idea is to avoid the worst. In this criterion the motive of self preservation is considered.

(ii) Maximax Criterion

If the decision maker is an optimist by nature, he would always think that the state of nature would be the best from his point of view. He will find out the expected pay-off of all the strategies and pick up the strategy which gives the maximum pay-off out of all the strategies. He will always think that the state of nature would be favourable.

4. Minimax Regret Criterion

When the criterion is in terms of cost or regret then the decision maker would choose the strategy in which the maximum regret or the cost is the lowest. The regrets have to be calculated for each act with reference to the best pay-off of the various alternative acts.

5. Laplace Criterion

Under this criterion when under conditions of uncertainty there is complete ignorance about the probability of the occurrences of state of nature, it is assumed that the probability of the occurrence of each state of nature is the same. After this, the strategy which maximises the expected pay-off is chosen.

6. Subjective Expected Utility Criterion

Under this criterion not only the knowledge gathered from past experience but also the judgment of the decision maker is taken into account in assigning probabilities to the states of nature. In this criterion, therefore, the expected value will be calculated by taking into account the posterior probabilities in regard to the state of nature in place of prior probabilities given.

Q10. Explain briefly the decision making under Uncertainty.

Ans : **(Imp.)**

The various criterias, used for decision making under uncertainty conditions are as follows,

(a) Optimism Criterion

Optimism criterion is also known as maximax or minimin criterion. This criterion was recommended by Leonid Hurwicz. Optimism criterion is totally opposite to pessimism criterion. Optimistic decision makers always try to maximize profit or minimize cost. In case of profit, optimism criterion aims at identifying the course of action or alternative strategy which denotes the maximum of maximum pay offs. In case of cost, optimism criteria aims at identifying the course of action or alternative strategy which denotes the minimum of minimum pay offs. As decision makers using this decision making criterion choose an alternative strategy or course of action with maximum (or minimum) possible pay off value, this criterion is also known as "optimistic decision criterion".

The steps involved in solving problem through optimism criterion are as follows,

1. Identify the best outcome from each course of action or alternative strategy. Maximum profit would be the best outcome in case of profit and minimum cost would be the best outcome in case of cost or loss.
2. Identify the highest (lowest) best outcome from the identified best outcomes of each course of action.
3. The alternative strategy or course of action corresponding to highest (lowest) best outcome is selected and declared as best strategy or optimal course of action.

(b) Realism Criterion

Criteria of Realism is also known as Hurwicz principle or hurwicz criteria which is named with the persons name Hurwicz. Hurwicz recommended Realism criterion. Hurwicz brought into notice the idea of coefficient of optimism which is represented by *a* (alpha), a is used to test the extent of optimism in the decision maker. Alpha(a) lies in between 0 and 1 (i.e., $0 < a < 1$). $a = 1$ denotes the optimistic anticipation of decision maker regarding the future and $a = 0$ denotes the pessimistic anticipation of decision maker regarding the future.

(c) Pessimism Criterion

Pessimism criteria is one of the criteria used for decision making under uncertainty situations. Pessimism criteria is also known as ' Waldian criterion' as this criteria was introduced by Abraham Wald. Pessimism criterion focuses on finding out the worst possible outcome of each course of action or strategy and selecting the best from the identified worst possible outcomes.

(d) Regret Criterion

Regret criterion of decision making was introduced by L.J. Savage. There are the chances wherein decision maker may face regret after making the decision. In order to avoid this regret feeling decision maker must make efforts to minimize regret before taking any decision or before choosing any alternative strategy.

(e) Equiprobable Criterion

'Laplace criterion' or 'Baye's criterion' was proposed by 'Thomas Bayes' and guided by 'Simar de Laplace'. Laplace criterion is also known as 'equal probabilities criterion' as this criterion is designed on an assumption that probabilities of all states of nature are equal.

Q11. Define Decision Tree Analysis. Explain various terms involved in Decision Tree Analysis.

Ans :

Meaning

A decision-tree is a schematic representation of a sequential and multidimensional decision problems. A decision-tree is made up of nodes, branches, probability estimates and payoffs. Nodes are basically of two types,

1. Decision Nodes

Decision nodes are generally represented by squares and depicts the places where a decision-maker takes a decision. The different possible courses of action available to the decision-maker are represented or indicated by each branch leading away from a node.

2. Chance Nodes

The chance nodes are represented by circles and shows a point at which the decision-maker will discover the response to the decision i.e., different possible outcomes. Which can be obtained from a selected course of action.

Branches coming out of the nodes connects several nodes which are either decisions or states of nature. Basically branches are of three types,

(i) Decision branches

(ii) Chance branches

(iii) Terminal branches.

3. Payoffs

The payoffs can be either positive i.e., revenue or sales or negative i.e., costs or expenditure and also they can be either associated with decision branches or the chance branches.

Q12. Write about the method and steps involved in construction of decision tree.

Ans :

Method

1. Identification of all the possible courses of action.

2. List the possible results i.e., 'states of nature' of each course of action specified.

3. Calculation of payoff of each possible combination of courses of action and results. Payoff will be in monetary terms usually.

4. Assigning probabilities to the different possible results for each given course of action. Likelihood of occurrence of a particular event is being indicated by the probability.

5. At last choose that course of action which gives the maximum payoff.

Steps

Decision tree is a diagrammatic representation of alternative courses of action and sequence of states of nature. Courses of action and states of nature (or outcomes) are arranged as branches of a tree.

The various steps involved in decision tree analysis are listed down below,

Step-1

Determine the number of decisions to be taken and the alternative strategies available for each decision in a sequential manner.

Step-2

Determine the outcome (or event) which may occur from each alternative strategy (course of action).

Step-3

Construct a tree diagram representing the order in which decisions are taken and outcomes are occurring. The decision tree diagram begins from left side and move towards right side.

Step-4

Determine the probabilities of occurrences of each state of nature.

Step-5

Determine the pay off values for each pair (or combination) of state of nature and course of action.

Step-6

Calculate expected pay off value for each course of action starting from right side of the decision tree.

Step-7

Select the course of action (or alternative strategy) with the best expected pay off value.

Step-8

Work backwards from last decision point to first decision point and at each decision point repeat the steps from step 4 to step 7.

PROBLEMS

6. A Finance Manager is considering drilling a well. In the past, only 70% of wells drilled were successful at 20 metres depth in that area. Moreover on finding no water at 20 metres, some persons in that area drilled it further upto 25 metres but 20% struck water at that level. The prevailing cost of drilling is ` 500 per metre. The Finance Manager estimated that in case he does not get water in his own well, he will have to pay ` 15,000 to buy water from outside for the same period of getting water from the well. The following decisions are considered:

(a) Do not drill any well.

(b) Drill upto 20 metres, and

(c) If no water is found at 20 metres, drill further upto 25 metres.

Draw an appropriate decision tree and determine the Financial Manager's optimal strategy.

Sol :

Based on the given information, the decision tree is shown in figure below:

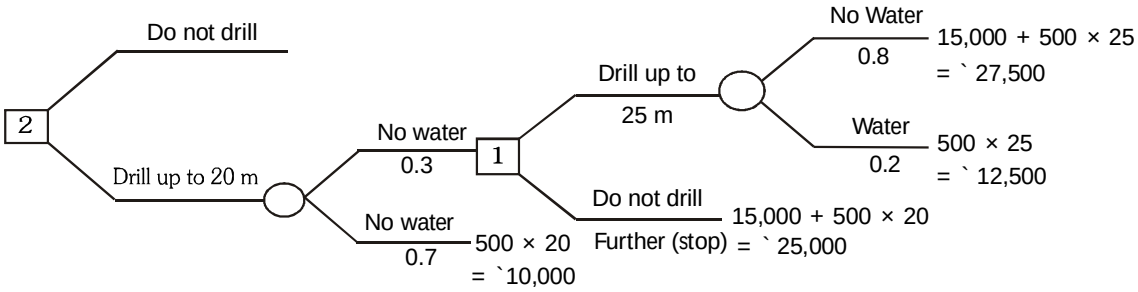


Fig.: Decision Tree Analysis Table

Analysis Table

Decision Node	Options	Expected Cost	Decision
2	Drill upto 25 metres	(0.8 × 27,500) + (0.2 × 12,500) = ` 24,500	Drill upto 25 metres
	Stop	` 25,000	
1	Do not drill	` 15,000	
	Drill upto 20 metres	(0.3 × 24,500) + (0.7 × 10,000) = ` 14,350	Drill upto 20 metres

From the analysis table, it is observed that at node 2 if it is decided to drill upto 20 metres and water is not found, then drill upto 25 metres. Node 1, involved lower expected cost, the decision is taken to drill upto 20 metres. Thus, the finance manager's optimal strategy is to drill upto 20 metres and if water is not struck then drill further to 25 metres.

5.5 GAME THEORY

5.5.1 Characteristics of Game Theory

Q13. Define game theory. Explain the characteristics and objectives of game theory.

Ans : **(Imp.)**

Meaning

In the competitive world, it is essential for an executive to study or at least guess the activities or actions of his competitor. Moreover, he has to plan his course of actions or reactions or counter actions when his competitor uses certain technique. Such war or game is a regular feature in the market and the competitors have to make their decisions in choosing their alternatives among the predicted outcomes so as to maximize the profits or minimizing the loss.

Characteristics

There can be various types of games. They can be classified on the basis of the following characteristics.

(i) Chance of Strategy

If in a game activities are determined by skill, it is said to be a game of strategy; if they are determined by change, it is a game of chance. In general, a game may involve game of strategy as well as a game of chance.

(ii) Number of Persons

A game is called an n-person game if the number of persons playing in n. The person means an individual or a group aiming at a particular objective.

(iii) Number of Activities

These may be finite or infinite.

(iv) Number of Alternatives (choices) Available to Each Person in a particular activity may also be finite or infinite. A finite game has a finite number of activities, each involving a finite number of alternatives, otherwise the game is said to be infinite.

(v) Information to the players about the past activities of other players is completely available, partly available, or not available at all.

(vi) Payoff

A quantitative measure of satisfaction a person gets at the end of each play is called a payoff. It is a real-valued function of variables in the game.

Objectives

(i) Refers to situation in which two or more players compete with each other.

(ii) Involves the players who have conflicting goals.

(iii) Deals with a situation in which there may be a number of possible outcomes with different values.

(iv) Has players going on strikes and hostels, some control that would influence the outcome but not having complete control over others.

Q14. State the basic terminology are used in game theory.

Ans :

1. Game Theory

A mathematical theory based on which strategy steps are employed to win a game played in a conflicting situation to maximise the benefits (or profit) or to minimise the damage (or loss).

2. Game

A competitive situation having the following characteristics:

(i) The situation is competitive.

(ii) There are a finite number of competitors (or players).

(iii) Each player has a finite number of strategies available to him or her.

(iv) The game is said to be played when both competitors initiate actions based on their chosen strategies.

(v) Every game results in an outcome.

(vi) Every outcome has stakes, i.e., payment given or taken.

3. Number of players

A game involving only two players (competitors) is called a two- person game. If the number of players exceeds two, then the game is known as "'n'- person game" where 'n' denotes the number of players.

4. Sum of gains and losses

If in a game the gains of one player are exactly the same as the losses to another player, such that the sum of the gains and losses equals zero, then the game is said to be a "zero-sum game". Otherwise it is said to be "non-zero sum game".

5. Strategy

It is a "plan of action" conceived and carefully executed by each party to the game. It involves a list of all possible actions (or moves or courses of action) that a player will take for every outcome (pay-off) that might arise. The rules governing the choices are known in advance to the players.

6. Optimal strategy

A particular strategy by which a player optimizes his gains or losses without knowing the competitor's strategies is called "optimal strategy".

7. Pure strategy

This is a predetermined plan of action based on which the games are played and which does not change during the game. It is a decision rule which is always used by the player to select the particular course of action.

8. Mixed strategy

It is a plan of action which is changed while the game is in progress, when both players are guessing as to which course of action to be selected on a particular occasion with some fixed probability.

9. Pay-off

The outcome of the game is known as "pay-off'.

10. Pay-off Matrix

A table (in the form of a matrix) showing the outcome of the game (in terms of gains or losses) when different strategies are adopted by the players.

11. Fair game

A game is said to be fair when the value of the game is zero.

12. Value of the game

The maximum guaranteed expected outcome per play when players follow their optimal strategy is called the "value of the game".

13. Solution of a game

When the best strategies of both players are determined and the value of the game is determined, we say, the game is solved or the solution of the game is obtained.

14. Maximin

The maximum value of the minimum pay-offs in each row.

15. Minimax

The minimum value of the maximum pay-offs in each column.

16. Saddle point

The game value is called the saddle point in which each player has a pure strategy. The saddle point is the lowest numerical value in a row and the largest numerical value in e column, which are equal to each other.

17. Strictly determinable game

A game is said to be strictly determinable if the maximin value is equal to the minimax value. In other words for the optimal strategy for both players, the pay-off for both players will be the same, i.e., the gain of one player equals the loss of another."

Q15. State the Assumptions of game Theory

Ans :

Assumptions

The underlying assumptions, the rules of the game as given as follows :

1. The player act rationally and intelligently.

2. Each player has available to him a finite set of possible courses of action.

3. The player attempt to maximize gains and minimize losses.

4. All relevant information is known to each players.

5. The players make individual decisions without direct communication.

6. The players simultaneously select their respective courses of action.

7. The pay off is fixed and determined in advance.

5.5.2 Limitations of Game Theory

Q16. Explain the advantages and disadvantages of game theory

Ans : **(Imp.)**

Advantages

1. Game theory keeps deep insight to few less known aspects, which arise in situations of conflicting interests.

2. Game theory creates a structure for analysis of decision-making in various situations like interdependence of firms etc.

3. For arriving at optimal strategy, game theory develops a scientific quantitative technique for two person zero-sum games.

Disadvantages

1. The highly unrealistic assumption of game theory is that the firm has prior knowledge about its competitor's strategy and is able to construct the payoff matrix for possible solutions, which is not correct. The main fact is that any firm is not exactly aware of its competitor's strategy. He can only make guesses about its strategy.

2. The hypothesis of maximin and minimax clearly shows that players are not risk lover and have whole knowledge about the strategies but the fact is that it is not possible.

3. It is totally impractical to understand that the several strategies followed by the rival player against others lead to an endless chain.

4. Most economic problems occur in the game if many players are involved in comparison to two-person constant sum game, which is not easy to

understand. For example, the number of sellers and buyers is quite large in monopolistic competition and the game theory does not provide any solution to it.

5. In real market situations, it is doubtful to find the use of mixed strategies for making non zero-sum games.

5.5.3 Two Persons Zero Sum Game

Q17. State the different types of games.

Ans :

1. Two-person games and n-person games

In two person games the players may have many possible choices open to them for each play of the game but the number of players remain only two. Hence it is called a two person game. In case of more than two persons, the game is generally called n-person game.

2. Zero Sum Game

A zero sum game is one in which the sum of the payment to all the competitors is zero for every possible outcome of the game is in a game if the sum of the points won equals the sum of the points lost.

3. Two person zero sum game

A game with two players, where the gain of one player equals the loss to the other is known as a two person Zero sum game. It is also called a rectangular game because their payoff matrix is in the rectangular form. The characteristics of such a game are

(i) Only two players participate in the game

(ii) Each player has a finite number of strategies to use

(iii) Each specific strategy results in a payoff

(iv) Total payoff to the two players at the end of each play is zero.

5.5.4 Maximum and Minimax Strategies

Q18. Explain Minimax and maximin principles used in solving games.

(OR)

Explain minimax and maximin principles used in solving games.

Ans :

It is evident from a two person zero sum game that the row player tries to maximize his minimum profit i.e., applying maximin principle and the column player tries to minimize his maximum losses i.e., applying minimax principle. In a payoff matrix minimum value in each row represents the minimum profit for a particular strategy. That means, the player will be confident that whatever may be the strategy of the other player, he may get the minimum gain. Therefore, from these rows, if the minimum values are calculated, then the player tries to maximize his least profit. This is referred to as maximin principle.

Similarly, the column player tries to minimize his maximum losses. The maximum value in each column is calculated and written beneath the table. The player then chooses the minimum value from these maximum values. This principle is called minimax principle.

If maximum value of row minima and the minimum value of column maxima is equal, then the game is said to have a saddle point.

5.5.5 Saddle Point

Q19. Explain briefly about saddle point.

Ans :

A saddle point is a position in the payoff matrix where the maximum of row minima coincides with the minimum of column maxima. The payoff at the saddle point is called the value of the game.

We shall denote the maximin value by γ, the minimax value of the game by $\overline{\gamma}$ and the value of the game by γ.

Note

(i) A game is said to be fair if

maximin value = minimax value = 0, i.e., if $\overline{\gamma} = \underline{\gamma} = 0$

(ii) A game is said to be strictly determinable if maximin value = minimax value $\neq 0$, $\underline{\gamma} = \gamma = \overline{\gamma}$.

5.5.6 Dominating Strategy

Q20. Explain briefly about dominance property.

Ans : **(Imp.)**

Sometimes it is observed that one of the pure strategies of either player is always inferior to atleast one of the remaining ones. The superior strategies are said to dominate the inferior ones. In such cases of dominance, we reduce the size of the payoff matrix by deleting those strategies which are dominated by others. The general rule for dominance are:

(i) If all the elements of a row, say K^{th} row, are less than or equal to the corresponding elements of any other row say r^{th} row, then K^{th} row is dominated by the r^{th} row.

(ii) If all the elements of a column, say K^{th} column, arc greater than or equal to the corresponding elements of any other column, say r^{th} column, then the K^{th} column is dominated by the r^{th} column.

(iii) Dominated rows and columns may be deleted to reduce the size of the pay-off matrix as the optimal strategies will remain unaffected.

(iv) If some linear combinations of some rows dominates i^{th} row, then the i^{th} row will be deleted. Similar arguments follow for column.

5.5.7 Mixed Strategy

Q21. Define Mixed Strategy. What are the various methods of Mixed Strategy?

Ans :

When there are no saddle points in the game, pure strategies as optimal strategies are not available. Such games can be resolved by using mixed strategy concept and following methods can be adopted:

1. Algebraic method

   Refer to Unit - V Q.No. 22

2. Graphical method

   Refer to Unit - V Q.No. 23

5.5.7.1 Algebric Method

Q22. Explain the concept of Algebric Method.

Ans : **(Imp.)**

Let us consider a 2×2 two-person zero sum game without any saddle point having the payoff matrix for player A.

$$\begin{array}{c} \\ A_1 \\ A_2 \end{array} \begin{array}{c} \begin{array}{cc} B_1 & B_{12} \end{array} \\ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \end{array}$$

The optimum mixed strategies

$$S_A = \begin{bmatrix} A_1 & A_2 \\ p_1 & p_2 \end{bmatrix} \text{ and } S_B = \begin{bmatrix} B_1 & B_2 \\ q_1 & q_2 \end{bmatrix}$$

where,

$$p_1 = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})},\ p_1 + p_2 = 1 \Rightarrow p_2 = 1 - p_1$$

$$q_1 = \frac{a_{22} - a_{12}}{(a_{11} + a_{22}) - (a_{12} + a_{21})},\ q_1 + q_2 = 1 \Rightarrow q_2 = 1 - q_1$$

The value of the game is $v = \frac{a_{11}a_{22} - a_{12}a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$

5.5.7.2 Graphical Method

Q23. Define graphical method. Write down the algorithm for 2 × n, m × 2 games.

Ans : **(Imp.)**

Meaning

If in a game one of the players have two strategies only then graphical method is used to solve this type of game. If the game has no saddle point and has a pay-off matrix of type n x 2 or 2 x n then this method is used.

Algorithm for 2 × n Games

Step 1:

With the help of dominance property, reduce the size of the pay-off matrix of Player A, if it is possible.

Step 2:

Let us consider that x is the probability of selection of Alternative 1 by Player A and 1 – x be the probability of selection of Alternative 2 by Player A. Find the expected gain function of Player A with respect to each of the alternatives of Player B.

Step 3:

Find the value of the gain from the gain functions which are derived in step 2, when x is equal to 0 as well as 1.

Step 4:

The gain functions are plotted on a graph by assuming a suitable scale. Keep x on X-axis and the gain on Y-axis.

Step 5:

Find the highest intersection point in the lower boundary of the graph because the Player A is a maximin player. Let it be the maximin point.

Step 6:

If only two lines are passing through the maximin point then it forms a 2 × 2 pay-off matrix from the original problem by retaining only the columns corresponding to those two lines and go to step 8; otherwise go to step 7.

Step 7:

Two lines are identified with opposite slopes passing through that point and form a 2 × 2 pay-off matrix from the original problem by retaining only the columns corresponding to those two lines which are having.

Step 8:

Solve this 2 × 2 game with the help of algebraic method.

Algorithm for m × 2 Games

Step 1:

With the help of dominance property reduce the size of the pay-off matrix of Player A, if it is possible.

Step 2:

Let us consider that y is the probability of selection of Alternative 1 by Player B and 1 – y be the probability of selection of Alternative 2 by Player B. Find the expected gain function of Player B with respect to each of the alternatives of Player A.

Step 3:

Find the value of the gain from the gain functions which are derived in step 2, when y is equal to 0 as well as 1.

Step 4:

The gain functions are plotted on a graph by assuming a suitable scale. Keep y on X-axis and the gain on Y-axis.

Step 5:

Find the lowest intersection point in the upper boundary of the graph because the Player B is a minimax player. Let it be the minimax point.

Step 6:

If only two lines passing through the maximin point then it form a 2 × 2 pay-off matrix from the original problem by retaining only the rows corresponding to those two lines and go to step 8; otherwise go to step 7.

Step 7:

Two lines are identified with opposite slopes passing through that point and form a 2 × 2 pay-off matrix from the original problem by retaining only the rows corresponding to those two lines which are having opposite slopes.

PROBLEMS

7. Solve the following game by giving optimum strategies for each player and value of the game.

		Player B			
		B_1	B_2	B_3	B_4
Player A	A_1	3	3	1	10
	A_2	5	5	4	6
	A_3	4	–2	0	–5

Sol :

Maximin Value

Step-1 : Identify the minimum value in each row.

		Player B				Row Minima
		B_1	B_2	B_3	B_4	
Player A	A_1	3	3	1	10	1
	A_2	5	5	4	6	4
	A_3	4	–2	0	–5	–5

Step-2

Select the maximum of the minimum value of each row and enclose it in a rectangle.

		Player B				Row Minima
		B_1	B_2	B_3	B_4	
Player A	A_1	3	3	1	10	1
	A_2	5	5	[4]	6	[4] → Maximin value
	A_3	4	–2	0	–5	–5

Minimax Value

Step-3

Identify the maximum value in each column.

		Player B			
		B_1	B_2	B_3	B_4
Player A	A_1	3	3	1	10
	A_2	5	5	4	6
	A_3	4	–2	0	–5
Column Maxima		5	5	4	10

Step-4

Select the minimum of the maximum value of each column and enclose it in a circle.

		Player B			
		B_1	B_2	B_3	B_4
Player A	A_1	3	3	1	10
	A_2	5	5	(4)	6
	A_3	4	–2	0	–5
Column Maxima		5	5	(4)	(10)

Minimax value

Saddle Point

Mark the maximum and minimum values in the payoff matrix.

		Player B				
		B_1	B_2	B_3	B_4	
Player A	A_1	3	3	1	10	1
	A_2	5	5	(4)	6	[4] → Maximin value
	A_3	4	–2	0	–5	–5
		5	5	(4)	10	

Minimax value

As the maximin and minimax value of the game coincides, it has a saddle point which is equal to '4'

$\therefore$ Value of game = 4

Optimal Strategy for Players

Player A – A_2

Player B – B_3

8. For a game played between two players the payoff of the first player is given in the following table.

		Player 2			
		1	**2**	**3**	**4**
Player 1	**1**	**3**	**-3**	**-2**	**-4**
	2	**-4**	**-2**	**-1**	**1**
	3	**1**	**-1**	**2**	**0**

Solve the game using at least two criterions.

Sol :

Criterion-1 : Maximum Value

Step-1- Identify the minimum value in each row.

Strategy		Player 2				Row Minima
		1	2	3	4	
Player 1	1	3	–3	–2	–4	–4
	2	–4	–2	–1	1	–4
	3	1	–1	2	0	–1

Step-2

Select the maximum of the minimum value of each row and enclose it in a rectangle.

Strategy		Player 2				Row Minima
		1	2	3	4	
Player 1	1	3	–3	–2	–4	–4
	2	–4	–2	–1	1	–4
	3	1	–1	2	0	–1 → Maximin Value

Criterion-2: Minimax Value

Step-3 - Identify the maximum value in each column.

Strategy		Player 2			
		1	2	3	4
Player 1	1	3	–3	–2	–4
	2	–4	–2	–1	1
	3	1	–1	2	0
	Column Maxima	3	–1	2	1

Step-4 - Select the minimum of the maximum value of each column and enclose it in a circle.

Strategy		Player 2			
		1	2	3	4
Player 1	1	3	–3	–2	–4
	2	–4	–2	–1	1
	3	1	–1	2	0
	Column Maxima	3	(–1)	2	1

↓ Minimax Value

Saddle Point

Mark the maximin and minimax value in the pay off matrix.

Strategy		Player 2				
		1	2	3	4	
Player 1	1	3	–3	–2	–4	–4
	2	–4	–2	–1	1	–4
	3	1	–1	2	0	–1 → Maximin
		3	–1	2	1	

Minimax

As the maximin and minimax value of the game coincides, it has a saddle point which is equal to '–1'.

Optimal Strategy for Players

Player 1 - Strategy 3

Player 2 - Strategy 2

Value of the game = –1

9. Consider the game having the following payoff table.

Strategy		Player 2			
		1	2	3	4
Player 1	1	5	-7	-2	2
	2	-2	2	-5	5
	3	-2	5	-2	7

Determine the optimal strategy for each player by successively eliminating dominated strategy. What is the value of the game? Can we call the game as a fair game?

Sol :

Given pay - off matrix is,

Strategy		Player 2			
		1	2	3	4
Player 1	1	5	–7	–2	2
	2	–2	2	–5	5
	3	–2	5	–2	7

Step - 1

Check for the saddle point.

Strategy		Player 2				Row Minima
		1	2	3	4	
Player 1	1	5	–7	–2	2	–7
	2	–2	2	–5	5	–5
	3	–2	5	(2)	7	[–2] → Maximin
Column Maxima		5	5	(–2) ↓	7	

Minimax

The minimax and maximin values are same. Hence, saddle point exist at - 2.

Step-2

Applying dominance principle to obtain the optimal strategies.

(a) Row Dominance

Compare player 1 's row 1 and row 2:

5 > –2, –7< 2, – 2 > – 5, 2< 5,

Compare player I's row 2 and row 3:

–2 = –2, 2<5, 5<– 2, 5<7

Since player 1 row 2 is dominated by row 3, hence row 2 is deleted.

The reduced matrix is

Strategy		Player 2			
		1	2	3	4
Player 1	1	5	–7	–2	2
	3	–2	5	–2	7

(b) Column Dominance

Compare player 2's column 3 and column 4

– 2< 2, – 2<7

Since Column 3 is dominated by column 4, column 4 is deleted

Now the reduced Matrix is,

Strategy		Player 2		
		1	2	3
Player 1	1	5	–7	–2
	3	–2	5	–2

Compare player 2's column 1 and column 3

$5 > -2, -2 = -2$

Since column 1 is dominated by column 3,

Column 1 is deleted

Now the reduced 2×2 Matrix is,

Strategy		Player 2	
		2	3
Player 1	1	–7	–2
	3	5	–2

Step 3

As the matrix is 2 x 2 size, apply algebraic method to get the solution.

Strategy		Player 2	
		2	3
Player 1	1	–7	–2
	3	5	–2

$\Rightarrow$

Strategy		Player 2	
		2	3
Player 1	1	a_{11}	a_{12}
	3	a_{21}	a_{22}

Thus, $a_{11} = -7, a_{12} = -2$

$a_{21} = 5, \ a_{22} = -2$

1. Let, player 1 selects a_2 strategy with probability x and strategy a_3 with probability $1 - x$.
2. Let, player 2 selects b_2 strategy with probability and strategy b_3 with probability $1 - y$.

Value of player 1 (a_2)

$$x = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$$

$$= \frac{(-2) - (5)}{[(-7) + (-2)] - [(-2) + (5)]}$$

$$= \frac{-7}{(-9)(3)} = \frac{-7}{-12} = \frac{7}{12}$$

$a_3 = 1 - x$

$$= 1 - \frac{7}{12}$$

$$= \frac{5}{12}$$

Value of Player 2 (b_2),

$$y = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$$

$$= \frac{-2 - (-2)}{[(-7) + (-2)] - [(-2) + (5)]} = \frac{0}{-9 - 3} = 0$$

$b_3 = 1 - y$

$= 1 - 0$

$= 1$

Value of the game,

$$V = \frac{a_{11}.a_{22} - a_{12}.a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$$

$$= \frac{(-7 \times -2) - (-2 \times 5)}{[(-7) + (-2)] - [(-2) + (5)]}$$

$$= \frac{14 + 10}{(-9) - 3} = \frac{24}{-12} = -2$$

Optimal Strategy

Particulars	Strategy	Probability
For Player -1	a_2	7/12
	a_3	5/12
For Player-2	b_2	0
	b_3	1

$\therefore$ Value of game, V = – 2

Since value of game is not equal to 0, the game is not a fair game.

10. Solve the following game using graphical method:

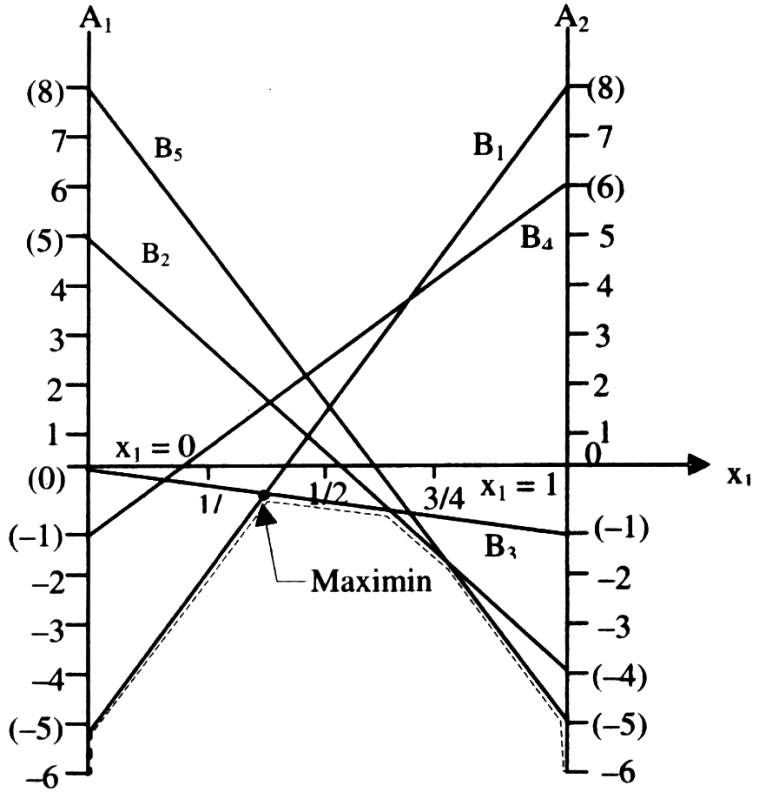
Particulars		Player B				
		B_1	B_2	B_3	B_4	B_5
Player A	A_1	– 5	5	0	–1	8
	A_2	8	– 4	–1	6	– 5

Sol :

After calculating the saddle point, it is clear that this does not contain any saddle point. Next, we have to reduce the size of the pay-off matrix using dominance rule. As the dominance rule cannot reduce the matrix, so let apply the graphical method. The player A's expected pay-offs along with the player B's pure strategies are shown in table below :

Player B's Pure Strategies	Player A's Expected Payoffs
1	$-5x_1 + 8(1-x_1) = -13x_1 + 8$
2	$5x_1 - 4(1-x_1) = 9x_1 - 4$
3	$0x_1 - 1(1-x_1) = x_1 - 1$
4	$-lx_1 + 6(1-x_1) = -7x_1 + 6$
5	$8x_1 - 5(1-x_1) = 13x_1 - 5$

Figure shows the graph for B's strategies where five lines are plotted as the functions of xi. The two lines (A] and A_2) parallel to each other are at a distance of one unit. Join the mark -5 located on line A] with mark 8 located on the line A_2 for representing B's first strategy. Similarly, joint mark 5 located on line Aj with mark -4 located on the line A_2 for representing B's second strategy and so on for other three strategies and bound the graph as shown in figure.



As the player A wants to maximise the minimum expected payoff, the two lines which are cutting at the upper point of the lower bound represent the two course of action B_1 and B_3 that player B select in his best strategy.

Thus one can reduce the 2×5 game matrix into 2×2 matrix that can be easily solved with the help of arithmetic method.

Now, the pay-off matrix of 2×2 game will be as follows :

Player A	B_1	B_2
	– 5	0
	8	– 1

This matrix will be of the following form:

$$\begin{matrix} & B_1 & B_3 \\ A_1 & a_{11} & a_{12} \\ A_2 & a_{21} & a_{22} \end{matrix}$$

The optimum mixed strategies for player A and B is illustrated by following matrices:

$$S_A = \begin{bmatrix} A_1 & A_2 \\ p_1 & p_2 \end{bmatrix}, p_1 + p_2 = 1$$

and $$S_B = \begin{bmatrix} B_1 & B_3 \\ q_1 & q_2 \end{bmatrix}, q_1 + q_2 = 1$$

Where,

$$p_1 = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}, p_1 + p_2 = 1 \Rightarrow p_2 = 1 - p_1$$

$$q_1 = \frac{a_{22} - a_{12}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}, q_1 + q_2 = 1 \Rightarrow q_2 = 1 - q_1$$

$$p_1 = \frac{-1-8}{(-5-1)-(0+8)} = \frac{-9}{-14} = \frac{9}{14}$$

$$p_2 = 1 - \frac{9}{14} = \frac{5}{14}$$

$$q_1 = \frac{-1-0}{(-5-1)-(0+8)} = \frac{-1}{-14} = \frac{1}{14}$$

$$q_2 = 1 - \frac{1}{14} = \frac{13}{14}$$

The optimum strategy of payoff matrix is now as follows :

$$S_A = \begin{bmatrix} A_1 & A_2 \\ \frac{9}{14} & \frac{5}{14} \end{bmatrix}, S_B = \begin{bmatrix} B_1 & B_3 \\ \frac{1}{14} & \frac{13}{14} \end{bmatrix}$$

The value of the game

$$(v) = \frac{a_{11}a_{22} - a_{12}a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$$

$$= \frac{(-5 \times -1) - (0 \times 8)}{(-5-1)-(0+8)} = -\frac{5}{14}$$

11. Solve the following game using graphical method:

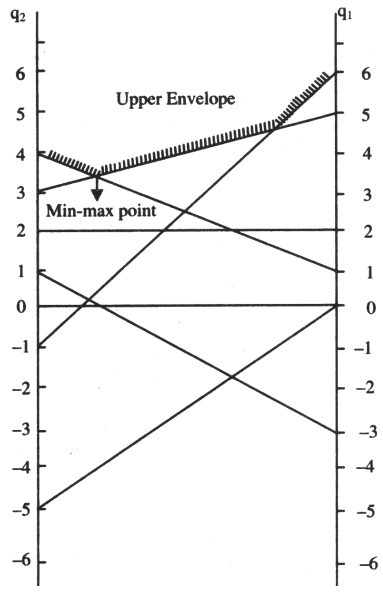
	Player B	
Player A	– 3	1
	5	3
	6	– 1
	1	4
	2	2
	0	– 5

Sol :

As this problem does not contain any saddle point, we have to apply the mixed strategy. Let assume that player B's mix strategy is given by $S_B = \begin{bmatrix} B_1 & B_2 \\ q_1 & q_2 \end{bmatrix}$ with $q_2 = 1 - q_1$ against player A.

The player B's mix strategy is illustrated in table below :

A's Pure Strategy	B's Expected Pay-Off
A_1	$-3q_1 + q_2$
A2	$5q_1 + 3q_2$
A3	$6q_1 - q_2$
A_4	$q_1 + 4q_2$
A5	$2q_1 + 2q_2$
As	$Oq_1 - 5q_2$



Now the reduced form of pay-off matrix will be as follows :

Player B

$$\text{Player A} \begin{matrix} \\ A_2 \\ A_4 \end{matrix} \begin{matrix} B_1 & B_2 \\ \end{matrix} \begin{bmatrix} 5 & 3 \\ 1 & 4 \end{bmatrix}$$

The optimum strategies are is given by the following matrices:

$$S_A = \begin{bmatrix} A_2 & A_4 \\ P_1 & P_2 \end{bmatrix}, S_B = \begin{bmatrix} B_1 & B_2 \\ q_1 & q_2 \end{bmatrix}$$

$$p_1 = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$$

$$= \frac{4-1}{(54)-(3+1)} = \frac{3}{5}$$

$$p_2 = 1 - p_1 = 1 - 3/5 = 2/5$$

$$q_1 = \frac{a_{22} - a_{12}}{(a_{11} + a_{22}) - (a_{12} + a_{21})} = \frac{4-3}{(5+4)-(3+1)} = \frac{1}{5}$$

$$q_2 = 1 - q_1 = 1/5 = 4/5$$

Value of the game $v = \frac{a_{11}a_{22} - a_{12}a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$

$$= \frac{5 \times 4 - 3 \times 1}{(5+4)-(3+1)} = \frac{17}{5}$$

5.6 LINEAR PROGRAMMING

5.6.1 Meaning

Q24. Define Linear Programming. Elucidate the Structure of LPP.

Ans : **(Imp.)**

Meaning

Linear programming deals with the optimization (maximization or minimization) of a function of variables known as objective functions. It is subject to a set of linear equalities and/or inequalities known as constraints. Linear programming is a mathematical technique which involves the allocation of limited resources in an optimal manner, on the basis of a given criterion of optimality.

To solve a LPP some basic requirements must be fulfilled. They are,

1. Objective function availability
2. Set of constraints
3. Both objective function and constraints must be linear functions. (or) Decision Variables.

Structure

The major components of LPP are,

1. Objective function
2. Constraints
3. Decision variables.

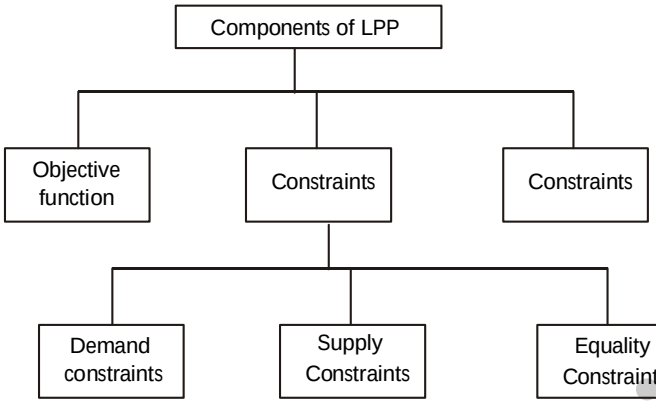


Fig.: Components of LPP

1. Objective Function

This is the function which is formulated by using contributions (i.e., profits, cost) and this function is to be optimized. That is to maximize the profit and minimize the loss respectively.

It is formulated by considering contributions and decision variables. It should be a linear equation.

2. Constraints

An objective function is always limited by some restrictions called as constraints.

(i) Demand Constraint

Demand is analogous to requirement. It has lower limit but upper limit is not mentioned.

(ii) Supply Constraint

Supply is related to availability. So it has upper limit.

(iii) Equality Constraint

Here the availability and requirement should be exactly equal.

3. Decision Variables

The variables which are required to be determined using different techniques are said to be decision variables.

Example

The variables in objective function x, *y* are nothing but the decision variables.

(a) Slack Variables

When a supply constraint is given i.e., '$\leq$' type of inequality. Thus, in order to change inequality to the equality a non-negative variable should be added on left side of an equation.

(b) Surplus Variables

When a requirement constraint is given i.e., '$\geq$' type of inequality. Thus in order to change inequality to equality, variable should be added to inequality on right hand side of an equation.

(c) Artificial Variables

The variable which should be added in case of minimization Linear Programming Problem (LPP) along with surplus variable is called an artificial variable.

5.6.2 Requirements for application

Q25. What are the Requirements for the Application of LPP?

Ans :

For a given problem situation, there are certain essential conditions that need to be solved by linear programming.

(i) Limited resources

Limited number of labour, material equipment and finance.

(ii) Objective

Refers to the aim to optimise (maximise the profit or minimise the costs).

(iii) Linearity

Increase in labour input will have a proportionate increase in output.

(iv) Homogeneity

The products, workers' efficiency, and machines are assumed to be identical.

(v) Divisibility

It is assumed that resources and products can be divided into fractions. (in case the fractions are not possible, like production of one-third of a computer, a modification of linear programming called integer programming can be used).

5.6.3 Assumptions

Q26. What are the Assumptions of LPP ?

(or)

Elucidate the various assumptions of LPP ?

Ans : **(Imp.)**

Following are the assumptions in linear programming problem that limit its applicability.

(a) Proportionality

A primary requirement of linear programming problem is that the objective function and every constraint function must be linear. Roughly speaking, it simply means that if 1 kg of a product costs Rs. 2, then 10 kg will cost Rs. 20. If a steel mill can produce 200 tons in 1 hour, it can produce 1000 tons in 5 hours.

Intuitively, linearity implies that the product of variables such as $x_1 x_2$, powers of variables such as x_3^2, and combination of variables such as $a_1 x_1 + a_2 \log x_2$, are not allowed.

(b) Additivity

Additivity means if it takes t_1 hours on machine G to make product A and t_2 hours to make product B, then the time on machine G devoted to produce A and B both is $t_1 + t_2$, provided the time required to change the machine from product A to B is negligible.

The additivity may not hold, in general. If we mix several liquids of different chemical composition, then the total volume of the mixture may not be the sum of the volume of individual liquids.

(c) Multiplicativity : It requires:

(i) If it takes one hour to make a single item on a given machine, it will take 10 hours to make 10 such items; and

(ii) The total profit from selling a given number of units is the unit profit times the number of units sold.

(d) Divisibility

It means that the fractional levels of variables must be permissible besides integral values.

(e) Deterministic

All the parameters in the linear programming models are assumed to be known exactly. While in actual practice, production may depend upon chance also. Such type of problems, where some of the coefficients are not known, are discussed in the extension of sensitivity analysis known as parametric programming.

5.6.4 Advantages

Q27. Explain the advantages LPP.

Ans :

Advantages

1. LP makes logical thinking and provides better insight into business problems.

2. Manager can select the best solution with the help of LP by evaluating the cost and profit of various alternatives.

3. LP provides an information base for optimum allocation of scarce resources.

4. LP assists in making adjustments according to changing conditions.

5. LP helps in solving multi-dimensional problems.

5.6.5 Application of LP

Q28. Explain the Applications of LPP.

Ans :

1. Production Management

Linear programming can be applied in production management for determining product mix, product smoothing, assembly line balancing, production scheduling and blending problem.

2. Marketing Management

Linear programming can be applied in marketing area by helping in analyzing the effectiveness of advertising campaign, advertising media mix and the shortest route for travelling salesman.

3. Manpower Management

LP allows the human resource to analyses personal policy combinations in terms of their appropriateness for maintaining a steady-state flow of people into the firm and out of the firm.

4. Transportation Problem

LP helps in determining the optimum transportation schedule with minimum total transportation cost of moving goods from various origins to various destinations.

5. Assignment Problem

LP models give the best assignment schedule with minimum total cost of assignment of various resources to various activities on a one-to-one basis.

6. Military Applications

LP models can be used for various military operations like wars, terrorists and defense against communal clashes etc. The problem is to determine the optimum number of weapons to be used that minimizes the cost of operations.

7. Agricultural Applications

LP can be applied in agricultural planning for allocating the limited resources such as water, labour, working capital etc., so as to maximize the net income.

8. Facilities Location

LP determines the most economic and efficient manner of locating manufacturing plants and distribution centres for physical distribution of goods.

***9.* Hydel Power Generation**

In this problem, variations in storage of dams which generate power is determined so as to maximize the energy obtained from the entire system.

10. Portfolio Selection

LP gives the optimal solution to the selection of specific investments from among a wide variety of alternatives. This helps the managers of banks, insurance companies, investment services etc.

11. Financial Mix Strategy

Financial mix strategy problems can be solved using LP technique. This involves the selection of means for financing projects, production operations and various other activities.

12. Airline Routine

LP can be used to determine the most economic pattern and timing for flights with the objective of efficiently using aircrafts, crews and money.

13. Diet Problems

LP helps the hospitals in determining the most economical diet for patients to meet the nutrients requirements of patients.

14. Environmental Protection

LP may be used to analyze alternatives for handling liquid waste material in order to satisfy antipollution requirements of patients.

15. Urban Development

Of late, LP is being used to analyze public expenditure planning, school busing, drug control, garbage collection etc.

5.6.6 Formulation of LP Problems

Q29. How LPP can be formulated? Discuss the mathematical formulation of LPP.

Ans : **(Imp.)**

The various steps/guidelines involved in the formulation of LPP are as follows,

(a) Identification and Selection of Variables

The first step in the formulation of LPP is identification of the variables called decision variables or design variables. After the identification of variables, it is then expressed in the form of symbolic notation along with their units of measurement. These variables may be seen as number of units to be ordered, number of units to be sold, number of units to be manufactured etc.

(b) Construction of Objective Function

The second step in the formulation of LPP is the setting of goal or objective function. It may be in two forms i.e., maximization or minimization form. If the data reveals profit in the system, then the objective function will be in maximization form. For example, output, profit, sales etc., are to be maximized. If the data is in terms of cost, then the objective function will be written in the minimization form. For example, production cost, spoilage, overheads, loss etc.

(c) Developing the Constraints Equation

The third step in the formulation of LPP is the development of the constraints equation. Constraints are the restrictions applied to the objective function. These constraints may be demand, supply and equality constraints. Demand constraints are represented by ' > ' sign and is used when a minimum restriction is applied to the function.

Supply constraints are represented by ' < ' sign and issued when a maximum restriction is applied to the function.

Equality constraints are represented by ' = ' sign and is used when an exact value should be used.

(d) Deciding the Variable Status

The fourth step in the formulation of LPP is deciding the status of decision variable. As the variables are generally numbers, they may be positive, negative or zero. For most of the cases it will be positive. For example, allocation of material for transportation cannot be negative, it should be either zero or greater than that. But, in some cases, it may be positive or negative, like in inventory. Therefore, the condition of the variable can be non-negative or unrestricted.

Mathematical Formulation of LPP

The general formulation of the LPP can be stated as follows:

In order to find the values of n decision variables $x_1, x_2, \ldots x_n$ to maximize or minimize the objective function.

$$Z = c_1x_1 + c_2x_2 + ... + c_nx_n \qquad ... (1)$$

and also satisfy m-constraints

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}\ x_n (\leq = \geq) b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}\ x_n (\leq = \geq) b_2 \\ \vdots \\ a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}\ x_n (\leq = \geq) b_i \\ \vdots \\ am_1x_1 + a_{m2}x_2 + ... + a_{mn}\ x_n (\leq = \geq) b_m \end{array}\right\} \quad ... (2)$$

where constraints may be in the form of inequality $\leq$ or $\geq$ or even in the form an equation (=) and finally satisfy the non negative restrictions.

$x_1 \geq 0, x_2 \geq 0, ... x_n \geq 0.$

PROBLEMS

12. Write about the formulation of LPP in agricultural aspect.

A farmer has a 100 acre farm. He can sell all the tomatoes, lettuce, radishes he can raise. The price he can obtain is ` 1 per kg for tomatoes, ` 0.75 a head for lettuce and ` 2 per kg for radishes. The average yield per acre is 2000 kg of tomatoes, 3000 heads of lettuce and 1000 kg of radishes. Fertilizer is available at ` 0.5 per kg and the amount required per acre is 100 kg each for tomatoes and lettuce and 50 kg for radishes. Labour required for sowing, cultivating and harvesting per acre is 5 man-days for tomatoes and radishes and 6 man-days for lettuce. A total of 400 man-days of labour are available at ` 20 per man-day. Solve the LPP in order to maximize the farmer's total profit.

Sol : **(Imp.)**

Decision Variables

Let, x_1 denote the number of acres allotted for tomato cultivation, x_2 denote the number of acres allotted for lettuce cultivation and let x_3 denote the number of acres allotted for radish cultivation.

Objective Function

The objective is to maximize the farmer's profit. The farmer's total profit = Sale - Expenditure.

Total sale of farmer,

$= ` [1 \times 2,000\ x_1 + 0.75 \times 3,000\ x_2 + 2 \times 1,000\ x_3]$

$= ` 2,000x_1 + 2,250x_2 + 2,000x_3$

Fertilizer expenditure,

$= ` 0.50\ (l00x_1 + 100x_2 + 50x_3)$

$= ` (50x_1 + 50x_2 + 25x_3)$

Labour expenditure,

$= ` 20.00\ (5x_1 + 6x_2 + 5x_3)$

$= ` (100x_1 + 120x_2 + 100x_3)$

Total expenditure is,

$` (150x_1 + 170x_2 + 125x_3)$

The total profit,

` [Sale - expenditure]

$= ` [2,000x_1 + 2,250x_2 + 2,000x_3] - ` [150x_1 + 170x_2 + 125x_3]$

$= ` (1,850x_1 + 2,080x_2 + 1,875x_3)$

Therefore, the objective function is, Maximize, $z = 1,850x_1 + 2,080x_2 + 1,875x_3$ ... (1)

Constraints

Since, the total area is restricted to 100 acres and the total man-days labour is restricted to 400 man-days, the constraints are,

$x_1 + x_2 + x_3 \leq 100$

$5x_1 + 6x_2 + 5x_3 \leq 400$ ... (2)

Non-negativity Restrictions

$x_1, x_2, x_3 \geq 0$ ...(3)

The relations equations (1) and (2) give a complete linear programming model for the problem.

13. A company having a mechanical workshop has recently discontinued production of an unprofitable product. It has resulted in a considerable spare capacity. The company has decided to use this capacity to the maximum extent to produce three products which are profitable. The productivity coefficient in machine hours per unit and available machine time is given below:

Machine Type	Product 1	Product 2	Product 3	Time Available Machine Hours per week
Milling Machine	9	3	5	500
Lathe	5	4	0	350
Grinder	3	0	2	150

The sales department has indicated that the demand for Products 1 and 2 exceeds the maximum production rate whereas sales potential for Product 3 is 20 units per week. The profits for the three products have been estimated respectively as ` 3500, ` 1400 and ` 1750 for the three products. The company wants to decide the optimum level of production to maximize its profit.

Formulate this problem as a mathematical model.

Sol : **(Imp.)**

Let,

x_1 = "Number of units of product 1 produced.

x_2 = Number of units of product 2 produced.

x_3 = Number of units of product 3 produced.

The product 1 consumes 9 hours, product 2 consumes 3 hours and products 3 consumables 5 hours on milling machine and the availability of milling machine per week is 500 hours. Therefore, the constraints can be written as,

$9x_1 + 3x_2 + 5x_3 \leq 500$

Similarly for Lathe and Grinder,

$5x_1 + 4x_2 + 0x_3 \leq 350$

and $3x_1 + 0x_2 + 2x_3 \leq 150$

The sales potential for product 3 is 20 units per week, therefore the constraint can be written as,

$x_3 \leq 20$

The objective here is to maximize the profit. Let, Z be the total profit per week

Max $Z = 3500x_1 + 1400x_2 + 1750x_3$

Finally, the above model can be represented as,

Maximize $Z = 3500x_1 + 1400x_2 + 1750x_3$

Subject to, $9x_1 + 3x_2 + 5x_3 \leq 500$

$5x_1 + 4x_2 \leq 350$

$3x_1 + 2x_3 \leq 150$

$x_3 \leq 20$

and $x_1, x_2, x_3 \geq 0$.

5.7 Graphical Solutions of LP Problems with Two Variables only

Q30. What is Graphical Method LPP ? Explain the characteristics of Graphical Method.

Ans :

Meaning

Graphical method is a simple method to understand and also to use. This is effectively used in LPP's which involves only 2 variables. It gives the graphical representation of the solutions.

All types of solutions are highlighted in this method very clearly. The only drawback is that more the number of constraints, more will be the straight lines which makes the graph difficult to understand.

Characteristics

The following are the characteristics of graphical method of LPP,

1. Method is very simple and easy to understand.
2. Very sensitive analysis and can be illustrated very easily by drawing graphs.
3. Very easy to obtain optimal solution.
4. It consumes very less time.

Q31. Describe the steps involved in graphical solution to linear programming models.

Ans : **(Imp.)**

Simple linear programming problems of two decision variables can be easily solved by graphical method. The outlines of graphical procedure are as follows :

Step 1 : Consider each inequality-constraint as equation.

Step 2 : Plot each equation on the graph, as each one will geometrically represent a straight line.

Step 3 : Shade the feasible region. Every point on the line will satisfy the equation of the line. If the inequality-constraint corresponding to that lines is '$\leq$', then the region below the line lying in the first quadrant (due to non-negativity of variables) is shaded. For the inequality-constraint with '$\geq$' sign, the region above the line in the first quadrant is shaded. The points lying in common region will satisfy all the constraints simultaneously. The common region thus obtained is called the feasible region.

Step 4 : Choose the convenient value of z (say = 0) and plot the objective function line.

Step 5 : Pull the objective function line until the extreme points of the feasible region. In the maximization case, this line will stop farthest from the origin and passing through at least one corner of the feasible region. In the minimization case, this line will stop nearest to the origin and passing through at least one corner of the feasible region.

Step 6 : Read the coordinates of the extreme point(s) selected in *step 5* and find the maximum or minimum (as the case may be) value of z.

Problems

14. A company produces two types of pens A and B. Pen A is of superior quality and pen B of inferior quality. Profit on pen A and pen B are ` 5 and ` 3 per pen respectively. Raw material required for each pen A is twice as that of pen B. The supply of raw material is sufficient only for 1000 pens of B per day. Pen A requires a special clip and only 400 clips are available per day. For pen B only 700 clips are available per day. Find graphically the product mix so that the company can make maximum profit.

Sol : **(Imp.)**

Step-1

Let x_1 = Number of Type A pens

x_2 = Number of Type B pens

The mathematical formulation of the problem is,

$$\text{Max } Z = 5x_1 + 3x_2$$

subject to $2x_1 + x_2 \leq 1000$

$x_1 \leq 400$

$x_2 \leq 700$

$x_1, x_2 \geq 0$

Step-2

Converting inequality constraints into equality constraints,

$2x_1 + x_2 = 1000$ ... (a)

$x_1 = 400$ ... (b)

$x_2 = 700$ ... (c)

Solving equation (a)

Assume value of $x_1 = 0$ and substitute in equation (a).

$2x_1 + x_2 = 1000$

$2(0) + x_2 = 1000$

$\boxed{x_2 = 1000}$

Assume $x_2 = 0$, substitute in equation (a)

$2x_1 + x_2 = 1000$

$2x_1 + 0 = 1000$

$2x_1 = 1000$

$x_1 = \frac{1000}{2} = 500$

$\boxed{x_1 = 500}$

$A = (x_1, x_2) = (500, 1000)$

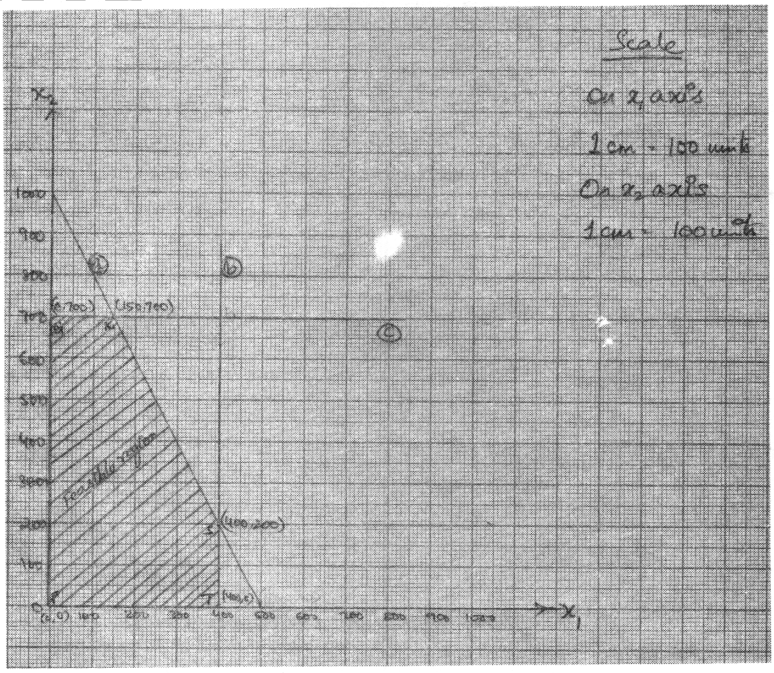
Equation (b) and (c) has only one variable, so no need to solve the equations.

$B = (x_1, x_2) = (400, 0)$

$C = (x_1, x_2) = (0, 700)$

Step-3

Plot values of A, B and C on graph to obtain feasible region.



Step-4

The coordinates of the extreme points of the feasible region are: P(0,0), Q (0, 700), R (150,700), S(400, 200), T(400, 0).

Substitute these values in objective function.

Extreme Points	Coordinates	Objective Function $Z = 5x_1 + 3x_2$
P	(0, 0)	5(0) + 3(0) = 0
Q	(0, 700)	5(0) + 3(700) = 2,100
R	(150, 700)	5(150) + 3(700) = 2,850
S	(400, 200)	5(400) + 3(200) = 2,600
T	(400, 0)	5(400) + 3(0) = 2,000

Hence, maximum profit of ` 2,850 can be earned when product mix involves $x_1 = 150$ and $x_2 = 700$.

15. The following table summarizes the key facts about two products, A and B, and resource Q, R and S, required to produce them.

Resource	Resource Usage Per Unit Produced		Amount of Resource Available
	Product A	Product B	
Q	2	1	2
R	1	2	2
S	3	3	4
Profit per unit	3	2	

Formulate the model and solve it graphically.

Sol : **(Imp.)**

Step-1

Objective Function

Max Z = 3A + 2B

Subject to,

$2A + 1B \leq 2$

$1A + 2B \leq 2$

$3A + 3B \leq 4$

and $A > 0, B > 0$.

Step-2

Converting inequality constraints into equality constraints.

$2A + 1B = 2 \quad \rightarrow (1)$

$1A + 2B = 2 \quad \rightarrow (2)$

$3A + 3B = 4 \quad \rightarrow (3)$

Step-3: Solving Equations

Equation 1

By solving equation (1), we get the values P and Q.

Case-I

If A = 0, then by solving equation (1), we get,

$2A + 1B = 2$

$2(0) + 1B = 2$

$0 + 1B = 2$

$\boxed{B = 2}$

$\therefore$ (0, 2)

Case-II

If B = 0, then by solving equation (1), we get,

$2A + 1B = 2$

$2A + 1(0) = 2$

$2A + 0 = 2$

$A = \frac{2}{2} \Rightarrow 1$

$\boxed{A = 1}$

$\therefore$ (1,0)

Equation 2

By solving equation (2), we get the values of R and S.

Case - I

If A = 0, then by solving equation (2), we get,

$1A + 2B = 2$

$1(0) + 2B = 2$

$B = \frac{2}{2}$

$\boxed{B = 1}$

$\therefore$ (0, 1)

Case - II

Let B = 0, then by solving equation (2), we get,

$1A + 2B = 2$

$1A + 2(0) = 2$

$1A = 2$

$\boxed{A = 2}$

$\therefore$ (2, 0)

Equation 3

By solving equation (3), we get T& U points,

Case - I

If A = 0 then by solving equation (3), we get,

$3A + 3B = 4$

$3(0) + 3B = 4$

$$\boxed{B = \frac{4}{3}} \Rightarrow B = 1.33$$

$\therefore$ (0, 1.33)

Case-II

If B = 0, then by solving equation (3), we get,

$3A + 3B = 4$

$3A + 3(0) = 4$

$3A + 0 = 4$

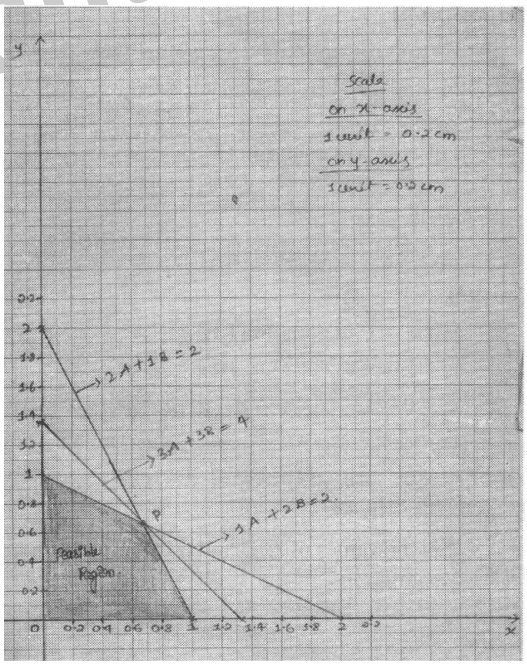
$$A = \frac{4}{3}$$

$$\boxed{A = 1.33}$$

$\therefore$ (1.33,0)

Step - 4

Plot the values obtained by solving three equations on the graph to find out the feasible region.



In the above graph it is seen that only the two equations i.e., 2A + 1B = 2 and 1A + 2B = 2 is effecting the feasible region. By solving these two points, we can determine point 'p'.

$2A + 1B = 2 \quad \rightarrow (4)$

$1A + 2B = 2 \quad \rightarrow (5)$

Multiply equation (4) with 's' and (5) with '2', we get,

$$\begin{array}{r} 2A + 1B = 2 \\ 2A + 4B = 4 \\ - \quad - \quad - \\ \hline -3B = -2 \end{array}$$

$$B = \frac{-2}{-3}$$

$$\boxed{B = \frac{2}{3}}$$

Substitute $B = \frac{2}{3}$ in equation (5), we get,

$1A + 2B = 2$

$$1A + 2\left(\frac{2}{3}\right) = 2$$

$$1A + \frac{4}{3} = 2$$

$$1A = 2 - \frac{4}{3}$$

$$1A = \frac{6-4}{3}$$

$$\boxed{A = \frac{2}{3}}$$

Point 'p' $= \left(\frac{2}{3}, \frac{2}{3}\right)$

Substitute these A and B values in objective function.

Max Z = 3A + 2B

$$= 3\left(\frac{2}{3}\right) + 2\left(\frac{2}{3}\right)$$

$$= \frac{6}{3} + \frac{4}{3}$$

$$= \frac{6+4}{3}$$

$$Z = \frac{10}{3}$$

$\therefore \quad \boxed{Z = 3.33}$

Short Question and Answers

1. Define EOL.

Ans :

Meaning

An alternative approach that refers to maximization of Expected Monetary Value (EMV) and minimization of Expected Opportunity Loss (EOL) is also known as expected value of regret. The difference between the highest profit for a state of nature and the actual profit which is obtained for the specific course of action is defined as EOL.

Hence, the amount of payoff that is lost due to the rejection of a course of action, which is having the greatest payoff for the state of nature that has actually appeared, is referred as EOL. That course of action is recommended for which EOL is minimum.

2. EVPI

Ans :

Meaning

Valuing information is very important especially when the information is purchased from a third party. Decision makers compute the expected profit with perfect information and compare it with the cost of acquiring information to see if the information is worth buying or not. Hence, expected value of perfect information is the maximum cost a decision maker is agreeable to pay to acquire the perfect information about the happening or non-happening of an event.

3. Define Decision Tree Analysis.

Ans :

Meaning

A decision-tree is a schematic representation of a sequential and multidimensional decision problems. A decision-tree is made up of nodes, branches, probability estimates and payoffs. Nodes are basically of two types,

1. Decision Nodes

Decision nodes are generally represented by squares and depicts the places where a decision-maker takes a decision. The different possible courses of action available to the decision-maker are represented or indicated by each branch leading away from a node.

2. Chance Nodes

The chance nodes are represented by circles and shows a point at which the decision-maker will discover the response to the decision i.e., different possible outcomes. Which can be obtained from a selected course of action.

Branches coming out of the nodes connects several nodes which are either decisions or states of nature. Basically branches are of three types,

(i) Decision branches

(ii) Chance branches

(iii) Terminal branches.

4 Define game theory.

Ans :

Meaning

In the competitive world, it is essential for an executive to study or at least guess the activities or actions of his competitor. Moreover, he has to plan his course of actions or reactions or counter actions when his competitor uses certain technique. Such war or game is a regular feature in the market and the competitors have to make their decisions in choosing their alternatives among the predicted outcomes so as to maximize the profits or minimizing the loss.

5. Assumptions of game Theory

Ans :

The underlying assumptions, the rules of the game as given as follows :

1. The player act rationally and intelligently.

2. Each player has available to him a finite set of possible courses of action.

3. The player attempt to maximize gains and minimize losses.

4. All relevant information is known to each players.

5. The players make individual decisions without direct communication.

6. The players simultaneously select their respective courses of action.

7. The pay off is fixed and determined in advance.

6. Disadvantages of game theory

Ans :

Disadvantages

1. The highly unrealistic assumption of game theory is that the firm has prior knowledge about its competitor's strategy and is able to construct the payoff matrix for possible solutions, which is not correct. The main fact is that any firm is not exactly aware of its competitor's strategy. He can only make guesses about its strategy.

2. The hypothesis of maximin and minimax clearly shows that players are not risk lover and have whole knowledge about the strategies but the fact is that it is not possible.

3. It is totally impractical to understand that the several strategies followed by the rival player against others lead to an endless chain.

7. Saddle point.

Ans :

A saddle point is a position in the payoff matrix where the maximum of row minima coincides with the minimum of column maxima. The payoff at the saddle point is called the value of the game.

We shall denote the maximin value by γ, the minimax value of the game by $\overline{\gamma}$ and the value of the game by γ.

Note

(i) A game is said to be fair if

maximin value = minimax value = 0, i.e., if $\overline{\gamma} = \underline{\gamma} = 0$

(ii) A game is said to be strictly determinable if maximin value = minimax value $\neq 0$, $\underline{\gamma} = \gamma = \overline{\gamma}$.

8. Dominance property.

Ans :

Sometimes it is observed that one of the pure strategies of either player is always inferior to atleast one of the remaining ones. The superior strategies are said to dominate the inferior ones. In such cases of dominance, we reduce the size of the payoff matrix by deleting those strategies which are dominated by others. The general rule for dominance are:

(i) If all the elements of a row, say K^{th} row, are less than or equal to the corresponding elements of any other row say r^{th} row, then K^{th} row is dominated by the r^{th} row.

(ii) If all the elements of a column, say K^{th} column, arc greater than or equal to the corresponding elements of any other column, say r^{th} column, then the K^{th} column is dominated by the r^{th} column.

(iii) Dominated rows and columns may be deleted to reduce the size of the pay-off matrix as the optimal strategies will remain unaffected.

(iv) If some linear combinations of some rows dominates i^{th} row, then the i^{th} row will be deleted. Similar arguments follow for column.

9. Define Linear Programming.

Ans :

Meaning

Linear programming deals with the optimization (maximization or minimization) of a function of variables known as objective functions. It is subject to a set of linear equalities and/or inequalities known as constraints. Linear programming is a mathematical technique which involves the allocation of limited resources in an optimal manner, on the basis of a given criterion of optimality.

10. What are the Assumptions of LPP ?

Ans :

Following are the assumptions in linear programming problem that limit its applicability.

(a) Proportionality

A primary requirement of linear programming problem is that the objective function and every constraint function must be linear. Roughly speaking, it simply means that if 1 kg of a product costs Rs. 2, then 10 kg will cost Rs. 20. If a steel mill can produce 200 tons in 1 hour, it can produce 1000 tons in 5 hours.

Intuitively, linearity implies that the product of variables such as $x_1 x_2$, powers of variables such as x_3^2, and combination of variables such as $a_1 x_1 + a_2 \log x_2$, are not allowed.

(b) Additivity

Additivity means if it takes t_1 hours on machine G to make product A and t_2 hours to make product B, then the time on machine G devoted to produce A and B both is $t_1 + t_2$, provided the time required to change the machine from product A to B is negligible.

The additivity may not hold, in general. If we mix several liquids of different chemical composition, then the total volume of the mixture may not be the sum of the volume of individual liquids.

(c) Multiplicativity : It requires:

(i) If it takes one hour to make a single item on a given machine, it will take 10 hours to make 10 such items; and

(ii) The total profit from selling a given number of units is the unit profit times the number of units sold.

(d) Divisibility

It means that the fractional levels of variables must be permissible besides integral values.

11. Advantages of Linear Programming.

Ans :

1. LP makes logical thinking and provides better insight into business problems.

2. Manager can select the best solution with the help of LP by evaluating the cost and profit of various alternatives.

3. LP provides an information base for optimum allocation of scarce resources.

4. LP assists in making adjustments according to changing conditions.

5. LP helps in solving multi-dimensional problems.

12. Applications of LPP.

Ans :

1. Production Management

Linear programming can be applied in production management for determining product mix, product smoothing, assembly line balancing, production scheduling and blending problem.

2. Marketing Management

Linear programming can be applied in marketing area by helping in analyzing the effectiveness of advertising campaign, advertising media mix and the shortest route for travelling salesman.

3. Manpower Management

LP allows the human resource to analyses personal policy combinations in terms of their appropriateness for maintaining a steady-state flow of people into the firm and out of the firm.

4. Transportation Problem

LP helps in determining the optimum transportation schedule with minimum total transportation cost of moving goods from various origins to various destinations.

5. Assignment Problem

LP models give the best assignment schedule with minimum total cost of assignment of various resources to various activities on a one-to-one basis.

Exercise Problems

1. Solve the following LP problem graphically :

Maximize $Z = 20X_1 + 80 X_2$

Subject to

$4X_1 + 6X_2 \leq 90$

$8X_1 + 6X_2 \leq 100$

$5X_1 + 4X_2 \leq 80$

X_1 and $X_2 \geq 0$

(Ans : $X_1 = 0$, $X_2 = 15$ and $Z = 1200$).

2. Solve the following LP problem using simplex method:

Maximize $Z = 3X_1 + 2X_2 + 5X_3$

Subject to

$X_1 + X_2 + X_3 \leq 9$

$2X_1 + 3X_2 + 5X_3 \leq 30$

$2X_1 - X_2 - X_3 \leq 8$

X_1, X_2 and $X_3 \geq 0$

(Ans : $X_1 = 5$, $X_2 = 0$, $X_3 = 4$ and $Z = 35$).

3. Mr. Sethi has ? 10,000 to invest in one of three options: A, B or C. The return on his investment depends on whether the economy experiences inflation, recession, or no change at all. The possible returns under each economic condition are given below,

Strategy	State of Nature		
	Inflation	Recession	No Change
A	2,000	1,200	1,500
B	3,000	800	1,000
C	2,500	1,000	1,800

What should he decide, using the pessimistic criterion, optimistic criterion, equally likely criterion and regret criterion?

(Ans: Choose A : ` 120, Choose B : ` 300, Choose C : ` 176.6, Choose C : ` 50).

4. You are given the following payoffs of three acts A_1, A_2 and A_3 and the events E_1, E_2, E_3.

States of Nature	Three Acts		
	A_1	A_2	A_3
E_1	25	-10	-125
E_2	400	440	400
E_3	650	740	750

The probabilities of the states of nature are 0.1, 0.7 and 0.2, respectively. Calculate and tabulate the EMV and conclude which would prove to be the best course of action.

(Ans: EMV (A_1) = 412.5, EMV(A_1) = 455, EMV(A_3) = 417.5).

---

5. Solve the following game and find its value

	B			
A	–5	2	0	7
	5	6	4	8
	4	0	2	–3

(Ans : A_2, B_3; 4).

---

6. Solve the game whose payoff matrix is

	B	
A	5	2
	3	4

by arithmetic method and verify the results by algebraic method. Calculate the game value.

(Ans : A(1/4, 3/4) B(1/2, 1/2); V = 3.5)

Choose the Correct Answers

1. The solution for an LPP with two exact constraints and no inequality constraints will be [d]
   (a) Infeasible (b) Unbounded
   (c) Multiple optimal (d) Unique point size solution
2. The feasible region will be _____ type if one of the constraints is exact type in a two variable LPP [b]
   (a) Point (b) Line
   (c) Area (d) Volume
3. In graphical solutions, if x is unrestricted and $y \geq 0$, we get the solution in _____ . [b]
   (a) 1st quadrant (b) 1st & 2nd quadrant
   (c) 1st & 4th quadrant (d) Any quadrant
4. In which of the following cases we do not maximise the objective function [d]
   (a) Sales (b) Profits
   (c) Contribution (d) Costs
5. If every point of a line drawn with any two points in feasible region fall in the feasible region the region is said to be [d]
   (a) Convex (b) Concave
   (c) Island (d) Infeasible
6. In an LPP, if the values given to the variable satisfy conditions and constraints but not the objective function, then the solution is [a]
   (a) Feasible solution (b) Optimal solution
   (c) Infeasible solution (d) We can not say
7. The feasible region in the form of a ring is _____. [b]
   (a) Convex (b) Concave
   (c) Concavo-convex (d) Convexo-concave
8. The number of basic feasible solutions in a feasible region will be [a]
   (a) Finite (b) Infinite
   (c) Zero (d) We can not say
9. Which of the following models does not use probabilities. [d]
   (a) Inventory models (b) Game theory models
   (c) Queuing models (d) Linear programming
10. Which of the following belongs to Operations Research model classified on the basis of time reference. [c]
   (a) Predictive model (b) Normative model
   (c) Dynamic model (d) Simulation

Fill in the blanks

1. The difference between the highest profit for a state of nature and the actual profit which is obtained for the specific course of action is defined as _____ .
2. The probabilities of an event before the collection of new information is known as _____ probabilities'.
3. Optimism criterion is also known as _____ or _____ criterion.
4. Criteria of Realism is also known as _____ principle
5. Regret criterion of decision making was introduced by _____ .
6. The outcome of the game is known as _____ .
7. The _____ point is the lowest numerical value in a row and the largest numerical value in e column, which are equal to each other.
8. The variables which are required to be determined using different techniques are said to be _____ variables.
9. A particular strategy by which a player optimizes his gains or losses without knowing the competitor's strategies is called _____ strategy".
10. The hypothesis testing is sometimes called _____ decision theory.

ANSWERS

1. EOL
2. Priori
3. Maximax, minimin
4. Hurwicz
5. L.J. Savage
6. Pay-off.
7. Saddle
8. Decision
9. Optimal
10. Classical

One Mark Answers

1. Payoff

Ans :

Payoffs are also known as conditional values or profits, and its effectiveness is linked with a particular combination of a plan of action and state of nature.

2. Payoff Table

Ans :

A payoff table is a record of the states of nature (outcomes), that is mutually exclusive and collectively exhaustive with a set of likely courses of action. The payoff is calculated for each combination of nature and course of action.

3. Courses of Action

Ans :

The courses of action or decision choices is also known as acts. Hence, for any difficult situation every possible plans of action must be included.

4. Opportunity Loss

Ans :

The failure of not choosing the most positive course of action or approach is responsible for incurring an opportunity loss.

5. Pessimism Criterion

Ans :

Pessimism criteria is one of the criteria used for decision making under uncertainty situations. Pessimism criteria is also known as ' Waldian criterion' as this criteria was introduced by Abraham Wald.

FACULTY OF COMMERCE

M.Com. III - Semester (CBCS) Examination

Model Paper - I

RESEARCH METHODOLOGY AND QUANTITATIVE TECHNIQUES

Time : 3 Hours] **[Max. Marks : 70**

Note : Answer all the questions in not more than one page each.

PART - A (5 × 4 = 20 Marks)

[Short Answer Type]

ANSWERS

1. Primary Data. **(Unit-I, SQA-4)**
2. Define Interpretation. **(Unit-II, SQA-3)**
3. Large sample. **(Unit-III, SQA-8)**
4. What are the Properties of Correlation? **(Unit-IV, SQA-6)**
5. What are the Assumptions of LPP ? **(Unit-V, SQA-10)**

PART - B (5 × 10 = 50 Marks)

[Essay Answer Type]

Note: Answer all the questions by using internal choice in not exceeding 4 pages each.

6. (a) Briefly describe the different steps involved in a Research Process. **(Unit-I, Q.No.4)**

OR

(b) Distinguish between nominal, ordinal, Interval and Ratio Scale. **(Unit-I, Q.No.16)**

7. (a) What is Interpretation of Data? What are the Essentials for Interpretation? **(Unit-II, No.1)**

OR

(b) Discuss briefly the stages in Report Preparation. **(Unit-II, Q.No.8)**

8. (a) Define the term Estimation. What are the different types of Estimations? **(Unit-III, Q.No.4)**

OR

(b) The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data

Item	1	2	3	4	5	6	7	8	9	10
Life in '000 hour	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average life time of bullbs is 4,000 hours.

(Unit-III, Prob.14)

9. (a) Enlist the techniques of analysis ot variance. Explain about briefly one way ANOVA. **(Unit-IV, Q.No.4)**

OR

(b) Following is the classification of 100 students according to their sex and height.

Test whether the height of the students is dependent upon the sex.

	Height	
Sex	Tall	Short
Male	30	40
Female	20	10

(Unit-V, Prob.13)

10. (a) A Finance Manager is considering drilling a well. In the past, only 70% of wells drilled were successful at 20 metres depth in that area. Moreover on finding no water at 20 metres, some persons in that area drilled it further upto 25 metres but 20% struck water at that level. The prevailing cost of drilling is ` 500 per metre. The Finance Manager estimated that in case he does not get water in his own well, he will have to pay ` 15,000 to buy water from outside for the same period of getting water from the well. The following decisions are considered:

(a) Do not drill any well.

(b) Drill upto 20 metres, and

(c) If no water is found at 20 metres, drill further upto 25 metres.

Draw an appropriate decision tree and determine the Financial Manager's optimal strategy.

(Unit-V, Prob.6)

OR

(b) The following table summarizes the key facts about two products, A and B, and resource Q, R and S, required to produce them.

Resource	Resource Usage Per Unit Produced		Amount of Resource Available
	Product A	Product B	
Q	2	1	2
R	1	2	2
S	3	3	4
Profit per unit	3	2	

Formulate the model and solve it graphically. **(Unit-I, Prob.15)**

FACULTY OF COMMERCE

M.Com. III - Semester (CBCS) Examination

Model Paper - II

RESEARCH METHODOLOGY AND QUANTITATIVE TECHNIQUES

Time : 3 Hours] **[Max. Marks : 70**

Note : Answer all the questions in not more than one page each.

PART - A (5 × 4 = 20 Marks)

[Short Answer Type]

ANSWERS

1. Secondary Data. **(Unit-I, SQA-5)**
2. Reports. **(Unit-II, SQA-2)**
3. Hypothesis. **(Unit-III, SQA-4)**
4. Association of Attributes. **(Unit-IV, SQA-3)**
5. Disadvantages of game theory. **(Unit-V, SQA-6)**

PART - B (5 × 10 = 50 Marks)

[Essay Answer Type]

Note: Answer all the questions by using internal choice in not exceeding 4 pages each.

6. (a) What is Research Plan? Discuss the components of a Research Plan. **(Unit-I, Q.No.6)**

OR

(b) What are the Guidelines / Precautions for Preparation of Questionnaire ? **(Unit-I, Q.No.20)**

7. (a) Explain the various methods of generalization. **(Unit-II, Q.No.4)**

OR

(b) Out line the structures of a Research Report. **(Unit-II, Q.No.10)**

8. (a) Define the term large sample. What are the assumptions of large samples. **(Unit-III, Q.No.17)**

OR

(b) Two types of drugs were used on 5 and 7 patients for reducing their weight. Drug A was imported and drug B indigenous. The decrease in the weight after using the drugs for six months was as follows :

Drug A	10	12	13	11	14		
Drug B	8	9	12	14	15	10	9

Is there a significant difference in the efficacy of the two drugs ? if not, which drug should you buy ?)For v = 10, $t_{0.05}$ = 2.223). **(Unit-III, Prob.15)**

9. (a) 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. **(Unit-IV, Prob.3)**

OR

(b) Four machines A, B, C, D are used to manufacture parts which are classified as first, second and third grade. Test whether the quality of the product is independent of the machines. The data as under,

Grade	A	B	C	D
First	620	750	400	530
Second	130	200	140	130
Third	50	50	60	40

(Unit-V, Prob.14)

10. (a) Solve the following game by giving optimum strategies for each player and value of the game.

		Player B			
		B_1	B_2	B_3	B_4
Player A	A_1	3	3	1	10
	A_2	5	5	4	6
	A_3	4	–2	0	–5

(Unit-V, Prob.7)

OR

(b) Define Linear Programming. Elucidate the Structure of LPP. **(Unit-I, Q.No.24)**

FACULTY OF COMMERCE

M.Com. III - Semester (CBCS) Examination

Model Paper - III

RESEARCH METHODOLOGY AND QUANTITATIVE TECHNIQUES

Time : 3 Hours] **[Max. Marks : 70**

Note : Answer all the questions in not more than one page each.

PART - A (5 × 4 = 20 Marks)

[Short Answer Type]

ANSWERS

1. What is Research Plan? **(Unit-I, SQA-3)**
2. Discuss the method of generalization. **(Unit-II, SQA-8)**
3. Hypothesis. **(Unit-III, SQA-1)**
4. Define F-test. **(Unit-IV, SQA-1)**
5. Define game theory. **(Unit-V, SQA-4)**

PART - B (5 × 10 = 50 Marks)

[Essay Answer Type]

Note: Answer all the questions by using internal choice in not exceeding 4 pages each.

6. (a) Define the term Research. Explain the purpose of Research. **(Unit-I, Q.No.1)**

OR

(b) Explain different types of Sampling Methods. **(Unit-I, Q.No.26)**

7. (a) Discuss how the statistical fallacies arise while interpreting the data. Explain with suitable examples. **(Unit-II, Q.No.5)**

OR

(b) Define and discuss footnotes in Research Report. Explain the purposes of foot notes. **(Unit-II, Q.No.11)**

8. (a) Explain the procedure and rules of testing hypothesis. **(Unit-III, Q.No.11)**

OR

(b) Eleven sales executive trainees are assigned selling jobs right after their recruitment. After a fortnight they are withdrawn from their field duties and given a month's training for executives sales. Sales executed by them in thousands of rupees before and after the training in the same period are listed below.

Sales (000 `) (Before training)	23	20	19	21	18	20	18	17	23	16	19
Sales (000 `) (After training)	24	19	21	18	20	22	20	20	23	20	27

Do these data indicate that the training has contributed to their performance? **(Unit-III, Prob.17)**

9. (a) To assess the significance of possible variation in performance in a certain test between the grammar schools of a city, a common test was given to a number of students taken at random. From the following results, make out an analysis of variance.

Schools			
A	B	C	D
8	12	18	13
10	11	12	9
12	9	16	12
8	14	6	16
7	4	8	15

To assess the significance of possible variationc in performance in a certain test between the grammar schools of a city is given below. From the following results make out an analysis of variance.

Schools			
A	B	C	D
8	12	18	13
10	11	12	9
12	9	16	12
8	14	6	16
7	4	8	15

(Unit-IV, Prob.2)

OR

(b) Five coins are tossed 3200 times and the number of heads appearing each time is noted. At the end, the following results were obtained.

No. of heads	0	1	2	3	4	5
Frequency	80	570	1100	900	500	50

Test the goodness of fit to determine whether the coins are unbiased. Use 5% LOS.

(Unit-V, Prob.15)

10. (a) Write about the formulation of LPP in agricultural aspect.

A farmer has a 100 acre farm. He can sell all the tomatoes, lettuce, radishes he can raise. The price he can obtain is ` 1 per kg for tomatoes, ` 0.75 a head for lettuce and ` 2 per kg for radishes. The average yield per acre is 2000 kg of tomatoes, 3000 heads of lettuce and 1000 kg of radishes. Fertilizer is available at ` 0.5 per kg and the amount required per acre is 100 kg each for tomatoes and lettuce and 50 kg for radishes. Labour required for sowing, cultivating and harvesting per acre is 5 man-days for tomatoes and radishes and 6 man-days for lettuce. A total of 400 man-days of labour are available at ` 20 per man-day. Solve the LPP in order to maximize the farmer's total profit. **(Unit-V, Q.No.12)**

OR

(b) Define graphical method. Write down the algorithm for 2 × n, m × 2 games. **(Unit-I, Q.No.23)**