



ALAGAPPA UNIVERSITY

[Accredited with 'A+' Grade by NAAC (CGPA:3.64) in the Third Cycle
and Graded as Category-I University by MHRD-UGC]

(A State University Established by the Government of Tamil Nadu)

KARAIKUDI – 630 003



Directorate of Distance Education

B.Com

VI - Semester

102 61

MARKETING RESEARCH

Reviewer	
Dr R Ganapathi	Assistant Professor, DDE, Alagappa University, Karaikudi

Authors

Dr Deepak Chawla, Distinguished Professor, Dean Research and Fellow Programme, International Management Institute (IMI), New Delhi

Dr Neena Sondhi, Professor, International Management Institute (IMI), New Delhi

Units: (1.2, 1.3, 2.3, 3.2, 3.3, 3.5.1, 4, 5, 6.2, 6.3, 6.5-6.6, 7.3-7.5, 8.3, 8.4, 8.6, 9.2-9.4, 10.5, 11.2-11.6, 11.8-11.9, 12.5)

S Sumathi, Marketing Consultant

P Saravanavel, Management Consultant

Units: (1.4, 1.5, 1.6, 3.4, 3.5, 6.2.1, 6.4, 7.2, 7.6, 8.2, 10.2-10.4, 12.2-12.4, 13.2, 13.3-13.5, 13.6-13.8, 14.2, 14.6.1)

Arun Kumar, Professor, New Delhi Institute of Management

N Meenakshi, Associate Professor, NIILM Centre for Management Studies

Units: (1.4.1, 1.7, 13.8.1, 14.3, 14.5)

Vikas® Publishing House, Units: (1.0, 1.1, 1.8-1.12, 2.0-2.2, 2.4-2.9, 3.0, 3.1, 3.6-3.10, 6.0, 6.1, 6.7-6.11, 7.0, 7.1, 7.7-7.11, 8.0, 8.1, 8.5, 8.7-8.11, 9.0, 9.1, 9.4.1, 9.5-9.9, 10.0, 10.1, 10.6-10.10, 11.0, 11.1, 11.7, 11.10-11.14, 12.0, 12.1, 12.6-12.10, 13.0, 13.1, 13.2.1, 13.5.1, 13.8.2, 13.9-13.13, 14.0, 14.1, 14.2.1-14.2.4, 14.4, 14.6, 14.7-14.11)

"The copyright shall be vested with Alagappa University"

All rights reserved. No part of this publication which is material protected by this copyright notice may be reproduced or transmitted or utilized or stored in any form or by any means now known or hereinafter invented, electronic, digital or mechanical, including photocopying, scanning, recording or by any information storage or retrieval system, without prior written permission from the Alagappa University, Karaikudi, Tamil Nadu.

Information contained in this book has been published by VIKAS® Publishing House Pvt. Ltd. and has been obtained by its Authors from sources believed to be reliable and are correct to the best of their knowledge. However, the Alagappa University, Publisher and its Authors shall in no event be liable for any errors, omissions or damages arising out of use of this information and specifically disclaim any implied warranties or merchantability or fitness for any particular use.



VIKAS®

Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.

E-28, Sector-8, Noida - 201301 (UP)

Phone: 0120-4078900 • Fax: 0120-4078999

Regd. Office: A-27, 2nd Floor, Mohan Co-operative Industrial Estate, New Delhi 1100 44

• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

Work Order No.AU/DDE/DE12-27/Preparation and Printing of Course Materials/2020 Dated 12.08.2020 Copies - 500

SYLLABI-BOOK MAPPING TABLE

Marketing Research

Syllabi	Mapping in Book
BLOCK I: BASICS OF MARKETING RESEARCH	
UNIT - I: Marketing Research Dynamics - Meaning of Research - Research Characteristics - Various Types of Research - Marketing Research and its Management - Nature and Scope of Marketing Research - Marketing Research in the 21st Century (Indian Scenario), Marketing Research - Value and Cost of Information.	Unit 1: Marketing Research Dynamics (Pages 1-25);
UNIT - II: Planning the Research Process - An Overview - Formulation of a Problem - Research Methods - Research Design- Data Collection Methods - Sample Design - Data Collection - Analysis and Interpretation - Report Writing - Components of Research Proposal - Components of Research Paper.	Unit 2: Planning the Research Process (Pages 26-38);
UNIT - III: Research Design - Meaning - Types of Research Design-Descriptive Research - Causal Research Design - Research Design and Marketing Decision Process - Choosing a Good Research Design.	Unit 3: Research Design (Pages 39-60);
UNIT - IV: Data Collection Sources and Methods - Meaning and Nature of Secondary Data - Advantages of Secondary Data - Drawbacks of Secondary Data - Types of Secondary Data Sources- Primary Data and its Types.	Unit 4: Data Collection Sources and Methods (Pages 61-81)
BLOCK II: RESEARCH METHODOLOGY	
UNIT - V: Designing Questionnaire - Questionnaire Building - Process of Questionnaire Design - Information Required - Interview Method - Questionnaire Format and Question Composition - Individual Question Content - Questions Order - Form and Layout- Pilot Testing the Questionnaire.	Unit 5: Designing Questionnaire (Pages 82-99);
UNIT - VI: Measurement and Scaling Techniques - Importance of Measurement and Scaling in Marketing Research - Scales of Measurement - Fundamental Properties - Primary Scales of Measurement - Attitude Measurement Scales - Types of Comparative Scales - Non - Comparative scale - Selecting an Appropriate Scale - Scale Evaluation.	Unit 6: Measurement and Scaling Techniques (Pages 100-118);
UNIT - VII: Sampling Theory-Design - size and Techniques - Meaning of Sampling - Importance of Sampling in Marketing Research - Sampling - Basic Constructs - Process of Sampling Design - Determining Sample Size - Probability and Non Probability Sampling - Classification of Probability Sampling Techniques - Classification of Non Probability Sampling - Selecting an Appropriate Sampling Technique.	Unit 7: Sampling Theory: Design, Size and Techniques (Pages 119-141);
UNIT - VIII: Hypothesis Testing - Hypothesis - Steps Involved in Hypothesis Testing - Test of Significance of a Mean - Test of Significance of Difference between Two Means - Test of Significance of Difference between Two Standard Deviations - Test of Significance of Sample Proportion - Test of Significance of Difference between Two Sample Proportions - Student's T - Distribution - Parametric and Non-Parametric Tests.	Unit 8: Hypothesis Testing (Pages 142-170)

BLOCK III: DATA ANALYSIS AND INTERPRETATION

UNIT - IX: Data Processing and Preliminary Data Analysis - Survey Field Work and Data Collection - Nature and Scope of Data Preparation - Editing, Coding - Data Entry - Data Cleaning - Preliminary Data Analysis - Assessing for Normality and Outliers.

UNIT - X: Data Analysis I - Statistics in Data Analysis - Measures of Central Tendency - Measures of Dispersion - Bivariate Analysis - Multivariate Analysis.

UNIT - XI: Data Analysis II - Descriptive Statistics - Univariate Analysis - Bivariate Analysis - Correlation analysis - analysis of variance and analysis of covariance - Multivariate Analysis - Conjoint analysis.

Unit 9: Data Processing and Preliminary Data Analysis
(Pages 171-184);

Unit 10: Data Analysis I
(Pages 185-192);

Unit 11: Data Analysis II
(Pages 193-260)

BLOCK IV: REPORT WRITING AND RECENT TREND IN MARKETING RESEARCH

UNIT - XII: Data Interpretation and Report Writing - Data Interpretation - Research Report - Modus Operandi of Writing a Market Research Report - Structure of the Report - Components of a Report - Style and Layout of a Report - Revising and Finalizing the Research Report - Quality Research Report - Responsibilities of a Market Research Report Writer - Presenting the Report.

UNIT - XIII: Applications of Marketing Research I - Consumer Market Research - Business-to-Business Market Research - Product Research - Pricing Research - Motivational Research - Distribution Research - Applications of Marketing Research II - Advertising Research - Media research - Sales Analysis and Forecasting - Data Mining.

UNIT - XIV: Recent Trends in Marketing Research - Marketing Information System and Research - Online Marketing Research - Recent Trends in Marketing Research - Research in Lifestyle Retail-Marketing Research and Social Marketing - Rural Marketing Research - Trends in Services Marketing Research - Brand Equity Research - International Marketing and Branding Research.

Unit 12: Data Interpretation and Report Writing
(Pages 261-282);

Unit 13: Application of Marketing Research
(Pages 283-317);

Unit 14: Recent Trends in Marketing Research
(Pages 318-342)

CONTENTS

INTRODUCTION

BLOCK I: BASICS OF MARKETING RESEARCH

UNIT 1 MARKETING RESEARCH DYNAMICS 1-25

- 1.0 Introduction
- 1.1 Objectives
- 1.2 Meaning of Research
 - 1.2.1 Research Characteristics
- 1.3 Various Types of Research
- 1.4 Marketing Research and its Management
 - 1.4.1 Management of Marketing Research
- 1.5 Nature and Scope of Marketing Research
- 1.6 Marketing Research in the 21st Century: Indian Scenario
- 1.7 Marketing Research: Value and Cost of Information
- 1.8 Answers to Check Your Progress Questions
- 1.9 Summary
- 1.10 Key Words
- 1.11 Self Assessment Questions and Exercises
- 1.12 Further Readings

UNIT 2 PLANNING THE RESEARCH PROCESS 26-38

- 2.0 Introduction
- 2.1 Objectives
- 2.2 Overview of the Research Process
- 2.3 Components of the Research Proposal
- 2.4 Components of a Research Paper
- 2.5 Answers to Check Your Progress Questions
- 2.6 Summary
- 2.7 Key Words
- 2.8 Self Assessment Questions and Exercises
- 2.9 Further Readings

UNIT 3 RESEARCH DESIGN 39-60

- 3.0 Introduction
- 3.1 Objectives
- 3.2 Meaning of Research Design
- 3.3 Types of Research Design: Descriptive Research and Causal Research Design
- 3.4 Research Design and Marketing Decision Process
- 3.5 Choosing a Good Research Design
 - 3.5.1 Selection of Research Design
- 3.6 Answers to Check Your Progress Questions
- 3.7 Summary
- 3.8 Key Words
- 3.9 Self Assessment Questions and Exercises
- 3.10 Further Readings

UNIT 4 DATA COLLECTION SOURCES AND METHODS 61-81

- 4.0 Introduction
- 4.1 Objectives
- 4.2 Meaning and Nature of Primary and Secondary Data
- 4.3 Uses, Advantages and Disadvantages of Secondary Data
- 4.4 Types of Secondary Data Sources
- 4.5 Primary Data and its Types
- 4.6 Answers to Check Your Progress Questions
- 4.7 Summary
- 4.8 Key Words
- 4.9 Self Assessment Questions and Exercises
- 4.10 Further Readings

BLOCK II: RESEARCH METHODOLOGY

UNIT 5 DESIGNING QUESTIONNAIRE 82-99

- 5.0 Introduction
- 5.1 Objectives
- 5.2 Questionnaire Building
- 5.3 Process of Questionnaire Design
- 5.4 Answers to Check Your Progress Questions
- 5.5 Summary
- 5.6 Key Words
- 5.7 Self Assessment Questions and Exercises
- 5.8 Further Readings

UNIT 6 MEASUREMENT AND SCALING TECHNIQUES 100-118

- 6.0 Introduction
- 6.1 Objectives
- 6.2 Scales of Measurement
 - 6.2.1 Importance of Measurement and Scaling in Marketing Research
- 6.3 Primary Scales of Measurement and their Fundamental Properties
- 6.4 Rating and Attitude Measurement Scales
- 6.5 Types of Comparative Scales and Non-Comparative Scales
- 6.6 Selecting an Appropriate Scale and Scale Evaluation
- 6.7 Answers to Check Your Progress Questions
- 6.8 Summary
- 6.9 Key Words
- 6.10 Self Assessment Questions and Exercises
- 6.11 Further Readings

UNIT 7 SAMPLING THEORY: DESIGN, SIZE AND TECHNIQUES 119-141

- 7.0 Introduction
- 7.1 Objectives
- 7.2 Meaning of Sampling
 - 7.2.1 Importance of Sampling in Marketing Research
- 7.3 Basic Constructs of Sampling
- 7.4 Determining the Sample Size
- 7.5 Process of Sampling Design: Probability and Non-Probability Sampling and their Classification

- 7.6 Selecting an Appropriate Sampling Technique
- 7.7 Answers to Check Your Progress Questions
- 7.8 Summary
- 7.9 Key Words
- 7.10 Self Assessment Questions and Exercises
- 7.11 Further Readings

UNIT 8 HYPOTHESIS TESTING

142-170

- 8.0 Introduction
- 8.1 Objectives
- 8.2 Hypothesis
- 8.3 Steps Involved in Hypothesis Testing
- 8.4 Parametric and Non-Parametric Tests
- 8.5 T-Distribution
- 8.6 Tests of Significance of a Mean, Standard Deviation and Proportion
 - 8.6.1 Test of Significance of Means - Case of Single Population
 - 8.6.2 Tests of Significance of Difference between two Population Means
 - 8.6.3 Test of Significance of Sample Population Proportion
 - 8.6.4 Tests Significance of Difference between two Population Proportions
- 8.7 Answers to Check Your Progress Questions
- 8.8 Summary
- 8.9 Key Words
- 8.10 Self Assessment Questions and Exercises
- 8.11 Further Readings

BLOCK II: DATA ANALYSIS AND INTERPRETATION

UNIT 9 DATA PROCESSING AND PRELIMINARY DATA ANALYSIS

171-184

- 9.0 Introduction
- 9.1 Objectives
- 9.2 Survey Field Work and Data Collection
- 9.3 Nature and Scope of Data Preparation: Editing and Coding
 - 9.3.1 Data Entry and Data Cleaning
- 9.4 Preliminary Data Analysis
 - 9.4.1 Assessing for Normality and Outliers
- 9.5 Answers to Check Your Progress Questions
- 9.6 Summary
- 9.7 Key Words
- 9.8 Self Assessment Questions and Exercises
- 9.9 Further Readings

UNIT 10 DATA ANALYSIS I

185-192

- 10.0 Introduction
- 10.1 Objectives
- 10.2 Statistics in Data Analysis
- 10.3 Measures of Central Tendency
- 10.4 Measures of Dispersion
- 10.5 Univariate, Bivariate and Multivariate Analysis of Data
- 10.6 Answers to Check Your Progress Questions

- 10.7 Summary
- 10.8 Key Words
- 10.9 Self Assessment Questions and Exercises
- 10.10 Further Readings

UNIT 11 DATA ANALYSIS II

193-260

- 11.0 Introduction
- 11.1 Objectives
- 11.2 Descriptive Statistics
- 11.3 Univariate Analysis
- 11.4 Bivariate Analysis
- 11.5 Correlation Analysis
- 11.6 Analysis of Variance
- 11.7 Analysis of Covariance
- 11.8 Multivariate Analysis
- 11.9 Conjoint Analysis
- 11.10 Answers to Check Your Progress Questions
- 11.11 Summary
- 11.12 Key Words
- 11.13 Self Assessment Questions and Exercises
- 11.14 Further Readings

BLOCK IV: REPORT WRITING AND RECENT TREND IN MARKETING RESEARCH

UNIT 12 DATA INTERPRETATION AND REPORT WRITING

261-282

- 12.0 Introduction
- 12.1 Objectives
- 12.2 Data Interpretation
- 12.3 Research Report
 - 12.3.1 Modus Operandi of Writing a Market Research Report
 - 12.3.2 Structure and Components of the Report
 - 12.3.3 Style, Layout of a Report and Finalizing the Report
 - 12.3.4 Quality Research Report
- 12.4 Responsibilities of a Market Research Report Writer
- 12.5 Presenting the Report
- 12.6 Answers to Check Your Progress Questions
- 12.7 Summary
- 12.8 Key Words
- 12.9 Self Assessment Questions and Exercises
- 12.10 Further Readings

UNIT 13 APPLICATION OF MARKETING RESEARCH

283-317

- 13.0 Introduction
- 13.1 Objectives
- 13.2 Consumer Market Research
 - 13.2.1 Business to Business Market Research
- 13.3 Product Research
- 13.4 Pricing Research

- 13.5 Motivational Research
 - 13.5.1 Distribution Research
- 13.6 Advertising Research
- 13.7 Media Research
- 13.8 Sales Analysis and Forecasting
 - 13.8.1 Forecasting Methods
 - 13.8.2 Dating Mining
- 13.9 Answers to Check Your Progress Questions
- 13.10 Summary
- 13.11 Key Words
- 13.12 Self Assessment Questions and Exercises
- 13.13 Further Readings

UNIT 14 RECENT TRENDS IN MARKETING RESEARCH

318-342

- 14.0 Introduction
- 14.1 Objectives
- 14.2 Marketing Information / System and Research
 - 14.2.1 Online Marketing Research
 - 14.2.2 Recent Trends in Marketing Research
 - 14.2.3 Research in Lifestyle Retail
 - 14.2.4 Marketing Research and Social Marketing
- 14.3 Rural Marketing Research
- 14.4 Trends in Services Marketing Research
- 14.5 Brand Equity Research
- 14.6 International Marketing Research
 - 14.6.1 Branding Research
- 14.7 Answers to Check Your Progress Questions
- 14.8 Summary
- 14.9 Key Words
- 14.10 Self Assessment Questions and Exercises
- 14.11 Further Readings

INTRODUCTION

NOTES

In business management one should pay attention of nine M's namely, money, materials, machines, men, method, market, minute, matter and milieu. Today management has recognised and realised the importance of "matter" i.e., 'information'. This is very much requisite in building up as well as sustaining a viable business in the rapidly changing market environment. In other words, the company works in a war zone of rapidly changing competitors, technological advances, new laws, trade policies and customer loyalty. Hence, it must urgently and critically think and rethink its business mission and marketing strategies for 'winning business'. The winning business is viewed as a 'system' for managing business process. To manage and master such business process majority depends on one measure called *Marketing Research*.

The marketing business is undergoing reappraisal in the light of the vast global, technological, economic and social challenges. The basic change, perhaps, in marketing thinking is a shift from pursuing a sale to creating customer. Transaction oriented, largely, has been replaced with more relationship-oriented designing the best marketing-mix for winning as well as keeping customer. Good customers are an asset to the company, which will in turn a handsome income stream to it. The underlying thrust of marketing research is envisioning new products, services, lifestyles and ways to raise living standards.

This book, *Marketing Research*, has been designed keeping in mind the self-instruction mode (SIM) format and follows a simple pattern, wherein each unit of the book begins with the Introduction followed by the Objectives for the topic. The content is then presented in a simple and easy-to-understand manner and is interspersed with Check Your Progress questions to reinforce the student's understanding of the topic. A list of Self-Assessment Questions and Exercises is also provided at the end of each unit. The Summary and Key Words further act as useful tools for students and are meant for effective recapitulation of the text.

BLOCK - I

BASICS OF MARKETING RESEARCH

*Marketing Research
Dynamics*

**UNIT 1 MARKETING RESEARCH
DYNAMICS**

NOTES

Structure

- 1.0 Introduction
- 1.1 Objectives
- 1.2 Meaning of Research
 - 1.2.1 Research Characteristics
- 1.3 Various Types of Research
- 1.4 Marketing Research and its Management
 - 1.4.1 Management of Marketing Research
- 1.5 Nature and Scope of Marketing Research
- 1.6 Marketing Research in the 21st Century: Indian Scenario
- 1.7 Marketing Research: Value and Cost of Information
- 1.8 Answers to Check Your Progress Questions
- 1.9 Summary
- 1.10 Key Words
- 1.11 Self Assessment Questions and Exercises
- 1.12 Further Readings

1.0 INTRODUCTION

Marketing Research is the process of gathering and analysing data in a structured and sequential manner regarding issues of marketing products and services provided by a business. The research so carried out must follow an interdisciplinary team approach, should use scientific methods that help in finding effective solutions and it should be economical too. There are several types of research that are carried out by businesses as per their needs and convenience. On the basis of nature of enquiry, there are two types of research namely exploratory research and conclusive research whereas on the basis of nature of investigation required, the types are descriptive research and causal research. This unit focuses on the dynamics and dimensions of marketing research.

1.1 OBJECTIVES

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

- Understand the meaning and types of Research
- Discuss the concept of Marketing Research and its management

*Self-Instructional
Material*

- Analyze the nature and scope of Marketing Research
- Describe the Indian scenario regarding Marketing Research
- Explain value and cost of information in Marketing Research

NOTES

1.2 MEANING OF RESEARCH

Different scholars have interpreted the term 'research' in many ways. For instance, Fred Kerlinger (1986) stated that 'Scientific research is a systematic, controlled and critical investigation of propositions about various phenomena.' Grinnell (1993) has simplified the debate and stated 'The word research is composed of two syllables, 're' and 'search'.

The dictionary defines the former as a prefix meaning 'again', 'anew' or 'over again'. Search is defined as a verb meaning 'to examine closely and carefully', 'to test and try', or 'to probe'. Together, they form a noun describing a careful, systematic, patient study and investigation in some field of knowledge, undertaken to establish facts or principles.'

Thus, drawing from the common threads of the above definitions, we derive that *management research is an unbiased, structured, and sequential method of enquiry, directed towards a clear implicit or explicit business objective*. This enquiry might lead to proving existing theorems and models or arriving at new theories and models. Let us now understand each part of the definition.

The most important and difficult task of a researcher is to be as objective and neutral as possible. Even though the researcher might have a lot of knowledge about the topic, he/she must not try to deliberately get results in the direction of the hypotheses.

The second thing to be remembered is that you follow a structured and sequential method of enquiry. For example, you may want to look at the available options to choose from if you want to study abroad. And you search the internet and ask your relatives and friends about the options for studying abroad. This is search and not research. For research, there must be a structured approach that you need to follow, and then only will it be called scientific. Thus, you may do a background analysis of how many students go abroad to study and based on this, form a hypotheses that 80 per cent of young Indians go to universities in the USA for further study. Then, you conduct a small survey amongst the students who are intending to go abroad for study. And based on the data collected, you are able to prove or disprove the hypotheses. So, we can state that you had conducted a research study. You will study the process of research later in the Unit.

The last and most important aspect of our definition that needs to be carefully considered is the decision-assisting nature of business research. As Easterby-

Smith, *et al.* (2002) state, business research must have some practical consequences, either immediately, when it is conducted for solving an immediate business problem or when the theory or model developed can be implemented and tested in a business setting. The world of business demands that managers and researchers work towards a goal—whether immediate or futuristic, else the research loses its significance in the field of management. The advantage with doing research is that one is able to take a decision with more confidence as one has tested it through research. For example, if you conduct a study of young women professionals and see that they have a need for a night crèche facility when they need to go out of town on official duty. You can conduct a small research to test what facilities they would like in this crèche and how much would they be willing to pay for this facility.

In fact, it would not be wrong to say that without the tool of research, there would be no new business practices or methods, as no one would want to start something new (for example, launch a new product, enter a new market segment, etc) without testing it through research.

1.2.1 Research Characteristics

The characteristics of research can be better understood by studying the features which made a research meaningful or good. For a research to be of value and to authenticate or contribute to the body of knowledge, we feel that it must possess the following characteristics:

- It must have a clearly stated purpose.
- It must follow a systematic and detailed plan for investigating the research problem. The source from which information is to be collected about quality standards inventory models has to be listed.
- The selection of techniques of collecting information, sampling plans and data analysis techniques must be supported by a logical justification.
- The results of the study must be presented in an unbiased, objective and neutral manner.
- The research that you undertake can never be fruitful if it corners or if it exploits the rights of the respondents. Thus, the research at every stage and at any cost must maintain the highest ethical standards.
- And lastly, the reason for a structured, ethical, justifiable and objective approach is the fact that the research carried out by us must be replicable. This means that the process followed by us must be 'reliable', i.e., in case the study is carried out under similar constraints and conditions it should be able to reveal similar results.

NOTES

NOTES

Check Your Progress

1. What is the most important and difficult task of a researcher?
2. What is the advantage of doing research?
3. What does the characteristic of replicability imply?

1.3 VARIOUS TYPES OF RESEARCH

Though every research conducted is unique, it is possible to categorize the research approach that you may decide to take. Figure 1.1 summarizes the types of research.

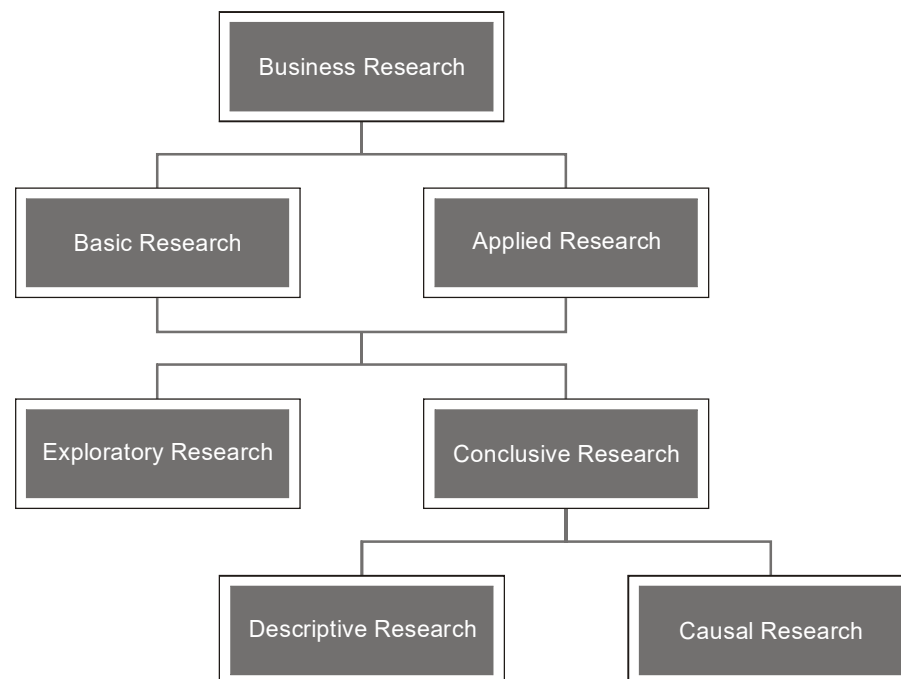


Fig. 1.1 Types of Research

Sometimes, research may be done for a purely academic reason of a need to know. For example, studies on employee dissatisfaction and attrition led to the study of impact of fixed working hours on family life and responsibilities. This study led to the organizations realizing that they need to have flexible working hours so that employees can better manage their work-life balance. The context of this kind of study is vast and time period, flexible. This type of research is termed as *fundamental* or *basic* research. On the other hand, you have studies that are specific to a particular business decision. For example, you find that despite being such an affordable car, the Tata Nano does not find a large number of buyers. Thus, the study you undertake would be of practical value to the specific organization. Secondly, it has implications for immediate action. This action-oriented research is termed as *applied* research.

NOTES

However, now we would like to advise you not to look at the two as opposites of each other. It may happen that the research which started as applied might lead to some fundamental and basic research, which expands the body of knowledge or vice versa. The process followed in both basic and applied research is systematic and scientific; the difference between them could simply be a matter of context and purpose.

Research studies can also be classified based on the nature of enquiry or objectives. Based on the nature of enquiry or objectives, research can be of the following types.

- Exploratory research
- Conclusive research

Exploratory Research

As the name suggests, exploratory research is used to gain a deeper understanding of the issue or problem that is troubling the decision maker. The idea is to provide direction to subsequent and more structured and rigorous research. The following are some examples of exploratory research:

- Let us say a diet food company wants to find out what kind of snacks customers like to eat and where they generally buy health food from.
- A reality show producer wants to make a show for children. He would like to know what kind of shows children like to watch.
- There is an investment bank that would like to know from its customers about what kind of help they want from the bank while making their investments.

As can be seen, for the examples above, an informal exploratory study would be needed. Exploratory research studies are less structured, more flexible in approach and sometimes could lead to some testable hypotheses. Exploratory studies are also conducted to develop the research questionnaire. (These will be discussed in detail in Unit 3.) The nature of the study being loosely structured means the researcher's skill in observing and recording all possible information will increase the accuracy of the findings.

Conclusive Research

Conclusive research is carried out to test and validate the study hypotheses. In contrast to exploratory research, these studies are more structured and definite. The variables and constructs in the research are clearly defined. For example, finding customer satisfaction levels of heavy consumers of different pizzas in the Pizza Hut menu. Now, this needs clear definition of customer satisfaction; secondly, how we will identify heavy consumers. The timeframe of the study and respondent selection are more formal and representative. The emphasis on reliability and validity of the research findings are all the more significant, as the results might need to be implemented.

NOTES

Based on the nature of investigation required, conclusive research can further be divided into the following types:

- Descriptive research
- Causal research

Descriptive Research

The main goal of descriptive research is to describe the data and characteristics about what is being studied. The annual census carried out by the Government of India is an example of descriptive research. The census describes the number of people living in a particular area. It also gives other related data about them. It is contemporary and time-bound. Some more examples of descriptive research are as follows:

- A study to distinguish between the characteristics of the customers who buy normal petrol and those who buy premium petrol.
- A study to find out the level of involvement of middle level versus senior level managers in a company's stock-related decisions
- A study on the organizational climate in different organizations.

All the above research studies are conducted to test specific hypotheses and trends. For example, we might hypothesize that the level of involvement of senior level managers is higher than middle level managers in stock-related decisions. They are more structured and require a formal, specific and systematic approach to sampling, collecting information and testing the data to verify the research hypotheses.

Causal Research

Causal research studies explore the effect of one thing on another and more specifically, the effect of one variable on another. For example, if a fast-food outlet currently sells vegetarian fare, what will be the impact on sales if the price of the vegetarian food is increased by 10 per cent. Causal research studies are highly structured and require a rigid sequential approach to sampling, data collection and data analysis. This kind of research, like research in pure sciences, requires experimentation to establish causality. In majority of the situations, it is quantitative in nature and requires statistical testing of the information collected.

Check Your Progress

4. Why is conclusive research carried out?
5. What is the main goal of descriptive research?
6. What do causal research studies explore?

1.4 MARKETING RESEARCH AND ITS MANAGEMENT

Marketing research can essentially be broken down into two distinct terms – *Marketing* and *Research*. We have to understand each of the two components *Marketing* and *Research* to know what Marketing Research – the combination of the two is all about.

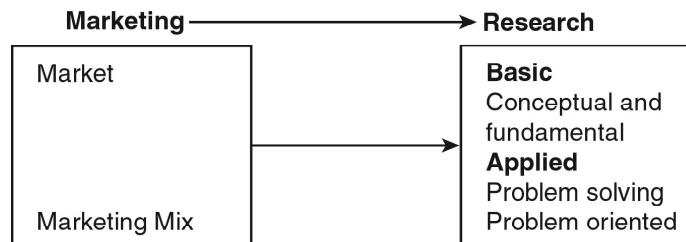


Fig. 1.2 Components of Marketing Research

Marketing is essentially a sum total of two entities, viz. the target market and the marketing mix for that market, which in turn is related to actions regarding the four Ps, viz., Product, Price, Place and Promotion. Figure 1.2 illustrates the two components of marketing research.

Thus, marketing is a human activity directed at satisfying needs and wants through an exchange process. A modern marketing system cannot exist without efficient producers and affluent consumers. The entire marketing decision-making environment is characterized by a combination of **controllable** and **uncontrollable factors**, i.e., causes and their effects determined in terms of behavioural responses and performance measures. The whole system showing the cause and effect relationship is explained in Figure 1.3.

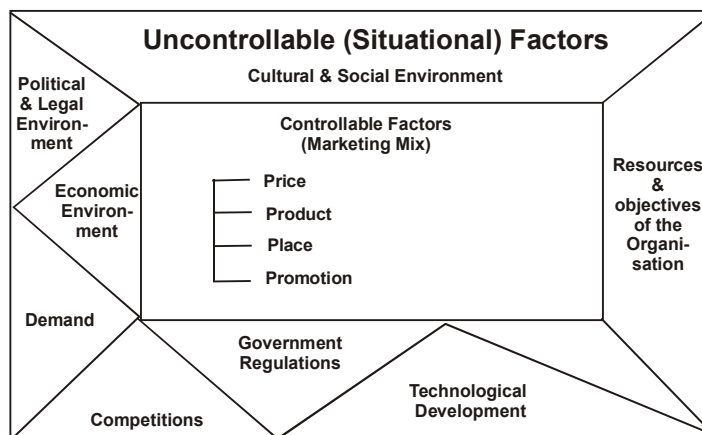


Fig. 1.3 Determinants of Marketing Decision-Making

Situational analysis is a creative process to understand the causal variables influencing the marketing system. Here diversity of information sources may be

NOTES

needed to develop insight and hypotheses regarding the causal factors. Flexibility in thinking and the use of multiple information sources are often critical to a successful situational analysis. Figure 1.4 illustrates certain behavioural responses.

NOTES

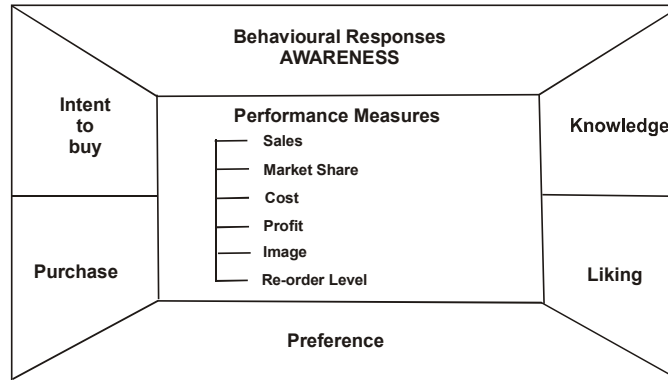


Fig. 1.4 Behavioural Responses

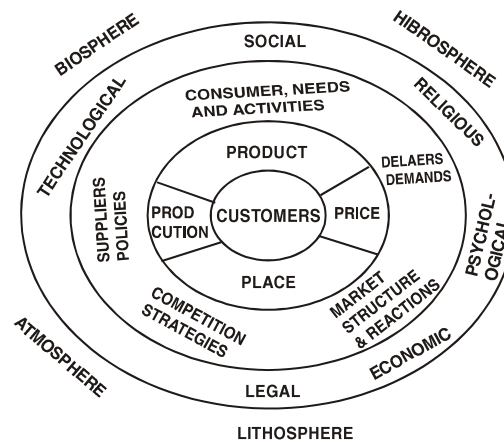


Fig. 1.5 Environmental Forces Influencing Marketing

Marketing management functions, viz., planning, implementation, and control, relating to marketing mix are also performed in an effective manner only when information on elements of marketing-mix and environmental forces (customers, needs and activities, market structure and reactions, competitors' strategies, dealers' demands, suppliers' policies, etc.,) are made available to marketing executives operating at various levels of management. Therefore, marketing research is a well-recognized and very important function of marketing management. Interactions of various environmental forces about which information is needed is shown diagrammatically in Figure 1.5.

Definition of Marketing Research

According to *Philip Kotler*, 'Marketing research is the systematic design, collection, analysis and reporting of data, and finding a solution relevant to specific marketing situation facing the company'. Marketing research is undertaken to

guide managers in their analysis, planning, implementation and control of programmes to satisfy customers and organizational goals.

*Marketing Research
Dynamics*

In the words of *Blankenship* and *Doyle*, marketing research is the collection and interpretation of facts that help marketing management to get products more efficiently into the hands of the consumer. Market research encompasses all information pertinent to this task, all the appropriate techniques.

This definition emphasizes the role of marketing research as a tool helpful to the marketing management. This, however, is a narrow definition as it does not specifically include services.

American Marketing Association (AMA) defines marketing research as ‘the systematic gathering, recording and analysis of data about problems relating to the marketing of goods and services.’ It may be noted that the definition (i) deals with a wider concept of marketing research as it includes all aspects of the marketing of goods or services and (ii) emphasizes that the data must be collected in a systematic manner to enhance objectivity. In fact, objectivity is the most important requisite for good research on any subject.

G. Zaltman and P.C. Burger have criticized the AMA’s definition of marketing research on the grounds that it does not include the pre-research analysis which is necessary to ascertain the type of information that should be collected, recorded and analyzed. He states,

Marketing research involves the diagnosis of information needs and the selection of relevant interrelated variables about which valid and reliable information is gathered, recorded, and analyzed.

It is evident that this definition emphasizes the diagnostic aspect of the information need. However, it does not explicitly mention the area of marketing within which the nature of information needs to be ascertained.

P. E. Green and D. S. Tull have defined marketing research as, ‘the systematic and objective search for and analysis of information relevant to the identification and solution of any problem in the field of marketing.’ This definition lays emphasis on the two requirements of good research, namely, systematic search and objectivity in the collection and analysis of data. The definition gives a wider ambit to marketing research.

1.4.1 Management of Marketing Research

Marketing research can be carried out in different ways, and the method the company chooses to use depends on the complexity of the problem it faces and the resources it is willing to expend.

- **The marketing department carries out the marketing research:** The company has a full fledged marketing department, but it does not expect it to handle extensive marketing research projects, and therefore it does not have a large marketing research budget.

NOTES

NOTES

Therefore, the marketing department carries out the marketing research when the study is small, like when data has to be collected from small number of respondents. This is usually true in business markets where the number of customers is small, or when the company is marketing a premium product whose clientele is very exclusive. This method allows company's marketers to understand customers' needs and behaviour in ways that cannot be replicated by studying reports of marketing research carried out by outside agencies. In fact, companies should mandate its marketers to be in constant interaction with its customers, and get a first hand feel of their aspirations and frustrations. This approach is also adopted when the information can be acquired from secondary sources like journals, internet, company records, etc.

- **The marketing research team carries out the marketing research:** The company hopes to carry out bigger marketing research projects, and hence creates a marketing research team within the marketing department itself. The marketing department briefs the team about the problem that it is expected to research, and the team designs and implements the project, and presents its finding to the marketing department. If the company feels that the marketing research project is too extensive to be handled by the marketing research team, the project is outsourced to a marketing research agency, and the team liaisons between the company and the agency.
- **The research project is designed by the marketing department, but the fieldwork is done by a marketing research agency:** The marketing department or the marketing research team designs the research project in terms of listing hypotheses that are to be tested, selection of sample and the survey methods to be used. The fieldwork, like conducting interviews, is done by the marketing research agency, and the results are again analysed by either the marketing department or the marketing research team.
- **A marketing research agency carries out the marketing research:** The marketing department believes that the research is too extensive to be handled by the marketing research team, or that the research project requires special competencies which it does not possess. Therefore, it assigns the research project to a marketing research agency. The marketing department holds detailed discussions with the marketing research agency, apprising it of the details of the problem it wants the agency to investigate. The agency designs the research project in terms of framing hypotheses, selecting sample, choosing survey methods and designing the questionnaire. It then conducts interviews, analyses the data, interprets the results, prepares a report and makes a presentation.

Check Your Progress

7. What is the marketing decision-making environment characterized by?
8. What, according to Philip Kotler, is marketing research?
9. Mention the role of marketing research agency in case they are hired for research.

NOTES

1.5 NATURE AND SCOPE OF MARKETING RESEARCH

In this section, you will study the nature and scope of marketing research. Objectives and importance of marketing research will be also discussed.

Nature of Marketing Research

1. **Marketing research is descriptive:** Marketing research describes the situation, the background of any marketing problem. It helps us to understand in a preliminary manner the underlying marketing problems.
2. **Marketing research is exploratory:** It explores the various decision alternatives, and examines them either to confirm / rule them out.
3. **Marketing research is evaluatory:** Marketing research compares two or more marketing strategies, for the selection of the best amongst them.
4. **Marketing research is predictive:** Marketing research prepares us for the future both in terms of risks and opportunities. Marketing research forecasts the future demand and market trends.
5. **Marketing research is an essential input:** Given its cost-effectiveness, it is significant for decision-making.
6. **Marketing is objective:** We do not pre-suppose things. We just put forth a hypothesis, collect data and either accept the hypothesis or reject it. While conducting problem-oriented research, we do not allow subjective bias to creep in. The research is never carried out to support a pre-determined result.
7. **Marketing research converts data into information:** Data, as such are not meaningful. However, on processing by systematic record, analysis and interpretation, data become useful information for the decision-maker.
8. **Marketing research is a continuous process:** Since marketing activity is a continuous one, where there are changes in product mix, promotion mix, distribution channels and prices, it is carried out on a regular basis. In the light of changed environment the data also changes, and a continuous monitoring is called for. These days marketing has become very dynamic,

NOTES

and market shares of different brands vary every month, requiring constant marketing research. In such cases, however, marketing research is integrated with Management Information Systems (MIS).

9. Marketing research draws heavily from different disciplines: By its very nature, marketing research draws heavily from Economics, Psychology, Sociology, Statistics, Operation Research, Information Technology and other similar subjects. Marketing research is a multi-dimensional, cross-disciplinary subject.

10. Marketing research is only a tool to the decision-maker: Marketing research does not supplant or replace the decision-making process. It only serves as an input to the decision-making process. It is at best a tool that guides the decision-maker. However, certain other variables may make the decision-maker take a decision contrary to the findings of marketing research, and he may want another marketing research study to be conducted for confirming or ruling out his decision.

Marketing research makes the total gamut of marketing activity more effective; it results into a right product mix with fair pricing, good distribution and effective promotion.

Some typical questions that a marketing research study would be able to answer may be as follows:

- Which segment of the society uses a particular brand as a routine procedure?
- What programmes boast of the highest TRP?
- Except money, what are the other factors that prevent people from travelling by air?
- Is the time right to launch a particular product in the market?
- What are the market requirements of a particular product?
- Is a product properly positioned? Does it compete effectively with other substitutes?
- What is the public opinion concerning public hospitals, airports and nationalized banks?
- What are the prospects of exporting handicraft items?
- Why do youngsters prefer romantic and/or thriller paperbacks?

These questions are related to marketing mix, sales, marketing environment, competition, market forecasting, buyer behaviour, etc. Marketing research does not answer technical/engineering questions like why a car is not fuel-efficient (however, it can answer questions like why people prefer a particular colour in a car).

Objectives of Marketing Research

Within the frame of research goals, namely academic (acquisition of knowledge) or utilitarian (immediate benefit), marketing research has the following objectives:

- To define the probable market for a product
- To know and report the market conditions and techniques, buying behaviour and buying power of customers
- To assess competitive strength of various marketing policies
- To forecast the market share and future sales
- To know the impact of promotional efforts
- To estimate potential buying power in various areas
- To indicate the distribution methods best suited to the product and the market
- To study the geographical distribution of a product
- To study the requirements regarding the features of a product
- To know about customer acceptance of a product
- To make packaging as attractive as possible
- To understand why consumers buy the product
- To identify marketing opportunities or problems
- To determine the dimension of the marketing problems
- To facilitate evaluation of the alternative solution of different problems
- To help in the selection of a right course of action

The information gathered by market research reduces the risk involved in marketing decisions. Marketing research is very useful in developing marketing strategies. It influences decisions on the pricing of the product and on channels of distribution, scale of advertising, etc. However, the information collected directly affects the planning of the product. For this reason, market research has been in great demand with firms making/marketing consumer goods, for example, confectionery, foodstuffs, cosmetics, soaps, etc. Marketing research is also used by firms producing branded goods and which are in competition with other brands. It helps them to know and maintain the popularity of their products among consumers.

Scope

The domain of marketing research has steadily been expanding its activities and techniques. The scope of marketing research covers the present and potential markets, marketing management and other areas of management. The coverage of marketing research is as follows:

1. **Market research:** It covers the different aspects, such as the size and nature of the market including export markets, dividing the consumers in

NOTES

NOTES

terms of their age, gender and income (market segmentation). It covers the economic aspects of marketing.

2. **Product research:** Product development and testing, brand name generation and testing, testing of existing products, packaging design studies, competitive product and product lines studies are covered within the field of product research.
3. **Service research:** The American Marketing Association (AMA) defines services as 'Activities, benefits of satisfactions which are offered for sale and are provided in connection with the sale of goods'. These services can be classified as: (a) Intelligible benefits like insurance, medical care, beauty care, etc. (b) Intangible activities which require the use of tangible goods like amusement, house rental, transport service, etc. (c) Intangible activities purchased jointly with goods, for example, credit facility, repairs, after sales services, etc.
4. **Price research:** Cost analysis, profit analysis, price elasticity, demand analysis, sales potential, sales forecasting, competitive pricing analysis, etc., are included within marketing research.
5. **Sales research:** It relates to the problem of regional variations in sales, fixing sales, priorities, measurement of the effectiveness of salesmen, evaluation of sales methods and incentives, etc.
6. **Distribution research:** Warehouse/plant location studies, channel performance coverage studies, export and international studies.
7. **Buyer/Consumer behaviour research:** This includes brand preference, attitude and awareness, product satisfaction, consumer behaviour, segmentation studies. Behavioural studies include questions like, Who is our target audience? What do they expect? How do the consumers behave? Why do they behave in a particular manner? All these and other related questions about consumers are answered by consumer research studies.
8. **Packaging research:** In essence, packaging research is a part of product research. However, recent developments in packaging and its contribution to advertising has led packaging to occupy an independent position. To know the impact and market response of packaging has become an independent research field.
9. **Promotion research:** Motivation research, media and advertising effectiveness studies, sales force compensation studies, etc., all cover the domain of promotion research.
10. **Brand image research:** Brand image, where it is a feature of advertising policy, particularly the consumer goods is said to establish a brand identity. It may begin with the name but may extend to other visual features like topography, colour, package, design, slogans, etc. These features should assist in creating, stimulating and maintaining demand.

NOTES

- 11. Corporate responsibility research:** Nowadays, environment-friendly organizations are fast replacing the traditional organizational set up. Studies assessing the environmental impact of organizations like ecological impact are studied along with studies on consumer rights, social values, social change, legal constraints studies, etc. These are becoming an important part of marketing research.
- 12. Business economics research:** Problems relating to input and analysis, forecasting, price and profit analysis, cost-volume-profit analysis (break-even charts), etc., are the main fields of this research.
- 13. Export marketing research:** This research studies the export potentials of the product. In such cases, any or all kinds of research mentioned above become necessary.
- 14. Qualitative research (QR):** Qualitative research has been accorded the respect it deserves in the recent past. The focus of qualitative research studies is on consumer products. Media research no longer merely involves numbers. Psychographic and life-style studies are common. Qualitative research is an understanding of the perceptions, attitudes and predispositions of a certain target segment towards a product, service, advertising concept or a need area. Qualitative research is based on an assumption that a common thread pervades through every segment of the consuming public. In a small sample of say four groups of eight respondents each is enough to appreciate this common thread. Qualitative research selects a few focus groups, e.g., say those who use conventional blades in a razor and those who use twin blades. It involves the study of what motivates these groups.
- 15. Miscellaneous research activities:** Various other research activities that are not conducted often in a company may be put in this category. This may include the research on diversification of products, motivation, satisfaction of sales personnel, assessment of the government's attitude towards corporate sector, research on international market potentialities, long-range demand forecasting, ecological impact studies, etc.

New Frontiers of Marketing Research

Until a few years ago, there was a reluctance to conduct syndicated studies. However, currently many syndicated studies are commissioned by multinational corporations. Marketing research is moving away from its problem-solving role and is playing a strategic role as research is being used from a strategic perspective. Growth areas include: (i) New product research, (ii) Financial research, (iii) Economic research, (iv) Service sector research, (v) Corporate image research, (vi) Industrial products research.

Distinction between Marketing Research and Market Research

Marketing research and *market research* are two separate terms. Some people use them interchangeably. Market research is a narrow term and is a part of

NOTES

marketing research. Market research strictly involves the study of the market itself to establish its size, growth or decline, the factors affecting it, the number of competitors active in it, their market share, the prices ruling in the market and so on. Thus, market research conveys a narrower meaning than marketing research, which covers the entire existing process and not merely the market. As a matter of fact, marketing research is not restricted to research on a single aspect of market, but it applies to any phase or any problem of marketing. It is the basis of sound decisions in the field of marketing. The exact difference between market research and marketing research has been explained by P. Kotler as, 'Some companies use the term market research instead by marketing research. The former term is accurate for describing research into markets—their size, geographical distribution, incomes and so on. However, it fails to cover the idea of research into the effects of marketing efforts on market, for which the term marketing research is more accurate.'

Importance of Marketing Research

Marketing research is the function that links the customer to the marketer through information gathered from various sources and methods. The information is used to identify and define marketing opportunities and problems; generate, refine and evaluate marketing actions; monitor marketing performance; and improve understanding of marketing as a process. Marketing research specifies the information required to address these issues, designs the method for collecting information, manages and implements the data collection process, analyzes the results, and communicates the findings and their implications.

Marketing research must provide accurate information that reflects a true state of affairs. It should be objective and should be conducted impartially. Although research is influenced by the researcher's philosophy, it should be free from personal or political biases of the researcher or the management. The researcher should never gather and analyze information to please or conform to the opinions of the management. Since the purpose of research is to act as an aid to decision-making for the management, the more unbiased and truthful such information is, the more accurate the decision would be. It is also important to remember that MR can only act as a tool for aiding decision-making. It can never provide the decision itself. The judgement of the manager, his experience and other factors would supplement the information gathered from marketing research.

Marketing information involves the identification, collection, analysis, dissemination and the use of information. Each phase of this process is important. The process starts with identifying or defining the MR problem or opportunity to determining what information is needed to investigate it, because every marketing opportunity translates into a research problem to be investigated, the terms 'objective' and 'problem' are used interchangeably. Next, the relevant sources of information are identified. A range of data collection methods varying in sophistication and complexity are evaluated for their usefulness and fit. Data are collected using the most appropriate method, they are analysed and interpreted,

and inferences are drawn. The findings and implications, and recommendations are provided to the management, which act on this information to make the correct decisions.

Check Your Progress

10. Which subjects does marketing research draw heavily from?
11. What does sales research relate to?
12. What is qualitative research?
13. What is the information derived in marketing research used for?

NOTES

1.6 MARKETING RESEARCH IN THE 21ST CENTURY: INDIAN SCENARIO

In India, there are some organizations that take up research studies on behalf of their clients. A complete list of such organization, the nature of research usually undertaken by them along with their strengths and limitations, is not yet available. As far as the syndicated services offered by Operations Research Group Indian Market Research Bureau, and Pathfinders India are concerned, ad hoc research projects are undertaken by research agencies at the instance of their clients. It may, however, be mentioned that these organizations (as well as some others) are well established and have been engaged in syndicated research and related spheres of activity for the past several years. They have a professionally qualified staff, and they undertake sponsored research on widely diversified themes.

An illustrative list of the customized research services offered by a leading research agency in India is given below:

Consumer Research

- Usage and attitude studies
- Brand image and positioning research
- New product development research
- Advertising research
- Brand tracking studies
- Product testing
- Simulated test marketing
- Market estimation and forecasting
- Market modelling
- Customized panel research

Qualitative Research

- Motivation research
- Life-style evaluation
- Concept evaluation
- Corporate image research
- Strategic research

The list is impressive as it shows a wide variety of research activities. Besides these two areas-consumer research and qualitative research, this research agency handles ad hoc research projects in other areas such as financial research, travel

NOTES

and tourism research, medical marketing research and social research. In a typical year it handles more than 400 widely diversified projects.

Another leading marketing research agency offers to undertake client-specific services in the following areas:

- Behaviour and attitude research
- Product and packaging studies
- Test marketing studies
- Corporate image studies
- Campaign evaluation
- Media studies
- Opinion surveys
- Industrial market research

It may be mentioned that in India, marketing research is still not well developed and there are only a few large marketing research agencies in the country. However, it goes to the credit of these agencies that some of the studies carried out by them were of pioneering nature, involving the development of appropriate concepts as well as sampling and measurement techniques. They have a large field setup supported by full-time investigators in different parts of the country having several years of professional expertise.

A company intending to buy marketing research may adopt one of the two courses. First, it may straightaway ask a marketing research agency to prepare a research proposal including an estimate of cost. Second, it may give some thought to its problem, spell it out and provide certain specifications according to which the concerned agency has to carry out the research. In this case the company will play the agency its fees for carrying out the task as per its specifications.

Marketing Research Scenario in Indian Corporate Sector

As long as marketers in India did not face much difficulty in selling their products (services) they did not find much value in marketing research. Formal use of marketing research in the Indian corporate sector began sometime in mid 1970's. In recent times, there is further momentum in the practice as the business environment in the country is now characterized by many diversities like:

- High degree of competition—the memories of long waiting time to get delivery of a product, limited choice and sometimes shabby products and unusually high prices have dimmed a great deal. The heat of competition is not restricted to the upper echelons of urban markets. Even the rural markets have begun to benefit from the impact of these changes.
- Large number of consumers with wide variations of requirements, tastes and preferences for different products and services.
- Change in the socio-economic condition of the market, specially the consumer goods manufacturers today encounter a sizeable middle income population, and moreover, rise in the number of working women and educational level have significantly affected consumer behavior.
- Growth in the communication facilities has led to greater awareness and more choice option to advertise through.

- Consumers, by and large, are found to have more discretionary income. But they have at the same time much concern about their value for money.
- Since 1992, liberal licensing and import policy of the Indian government have brought about many changes in the corporate environment. Changes have taken place both, in the scale and in the nature of problems which the management must tackle.

Recent developments in the Indian market bear a similarity to traditional free markets, where primary determinant of success is keeping the customers happy. Managing a product (service) in tune with such market forces assumes a critical role. Historically, in the shortage prone economy, marketing was found to be a luxury that did not seem relevant to the needs of the Indian consumer. This view is now rapidly changing. Expenditure on unsuccessful product launches, wrong product positioning, improper distribution system etc. can whittle away the vitals of an organization in the absence of a sound, professional marketing management and high quality marketing research. Thus corporate world in India has now realised the utility of data based marketing decision making rather than only intuition and hunch. So to derive an edge over the competition in today's dynamic business environment, any marketer finds it essential to evolve marketing programmes based on proper marketing research.

Check Your Progress

14. List the two courses which a company intending to buy marketing research may adopt.
15. Which factors can whittle away the vitals of an organization?

1.7 MARKETING RESEARCH: VALUE AND COST OF INFORMATION

Effective marketing starts with a strong knowledge of customers: the kind of knowledge that gives an organization unique insights into what customers want and how to satisfy them better than the competition. The most reliable source of fresh customer insights is good marketing information. Without good information, no marketing activity can succeed. Useful marketing information may come from a variety of sources both inside and outside the organization. Marketing information is generated by a variety of different activities, including marketing research.

Cost of Information

Before the advent of internet technologies, a marketer had to make an important trade-off about the type of information he could send and the media by which he could send it. If he wanted to send small messages to a wider audience, he could use advertising and if he wanted to send extensive messages to a small audience, he

NOTES

NOTES

could use salespersons to visit these customers and deliver the message. The marketer could send extensive messages to smaller audiences and he could also send brief messages to a large audience at a reasonable cost. But if he wanted to send extensive messages to a large audience, it was not possible to do so without enrolling an army of salesmen, and if he wanted to deliver a brief message to a small audience, it would be very costly both through advertising and employing salespersons.

This economics of information puts constraints on the development of the most efficient methods to deliver products and services. For instance, industrial buyers needed plenty of product and other information, and salespersons were employed to provide this information. But the industrial buyer had no interest in whether the product was delivered directly by the manufacturer or through a distributor. However, since the company was dealing directly with the customer, it also chose to deliver the product directly, which is always more costly than doing it through a distributor. In most cases, the choice of the information channel was determining the choice of the product channel. But product and information channels have different economics of operation (it is always economical to transfer products in bulk for maximum possible distance and then this bulk can be broken into smaller lots, if needed, to be transported over smaller distances to customers), and combined together to create an efficient system of transferring products and information.

Internet has changed the economics of information. Now it is possible to transfer brief as well as detailed messages to both small and wide audiences. The marketer does not make compromises between richness of message and its reach. Now he can get both. It is important that marketers understand the implication of the new economics of information and combine it judiciously with economics of movement of goods. Marketers will have to design separate channels for products and information and let their individual economics govern their configuration. For instance, if the product and other information can be provided by information on the seller's website, email and video-conferencing, the product can be made available through a distributor. Similarly, a retail store does the twin tasks of providing product information to the customer and stocking products. Providing information to customers and stocking products again have different economics of information and combining them together in one channel, in this case the retail store, is inefficient.

Marketers have made serious errors. While they understood and were eager to exploit the new economics of transfer of information, they forgot the economics of transfer of products. In fact they made the blunder of believing that products can be transferred as easily and as cheaply as information, and that transfer of an individual product via courier is same as bulk transfer. Marketers again decided the channel of products according to the channel of information they were using and incurred unsustainable costs.

Internet marketing will prosper as it is based on strong and irrefutable advantage of easy and cheap dispersion of all types of information to all types of audiences. But channels of information and product should be decided based on their individual economies. One should never be allowed to influence the other.

NOTES

Check Your Progress

16. What does effective marketing start with?
17. What did a marketer have to do before the advent of internet technologies?

1.8 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The most important and difficult task of a researcher is to be as objective and neutral as possible.
2. The advantage of doing research is that one is able to take a decision with more confidence as one has tested it through research.
3. The characteristic of replicability means that the process followed must be 'reliable', i.e., in case the study is carried out under similar constraints and conditions it should be able to reveal similar results.
4. Conclusive research is carried out to test and validate the study hypotheses.
5. The main goal of descriptive research is to describe the data and characteristics about what is being studied.
6. Causal research studies explore the effect of one thing on another and more specifically, the effect of one variable on another.
7. The entire marketing decision-making environment is characterized by a combination of *controllable* and *uncontrollable factors*, i.e., causes and their effects determined in terms of behavioural responses and performance measures.
8. According to *Philip Kotler*, 'Marketing research is the systematic design, collection, analysis and reporting of data, and finding a solution relevant to specific marketing situation facing the company'.
9. The agency designs the research project in terms of framing hypotheses, selecting sample, choosing survey methods and designing the questionnaire. It then conducts interviews, analyses the data, interprets the results, prepares a report and makes a presentation.
10. Marketing research draws heavily from Economics, Psychology, Sociology, Statistics, Operation Research, Information Technology and other similar subjects.

NOTES

11. Sales research relates to the problem of regional variations in sales, fixing sales, priorities, measurement of the effectiveness of salesmen, evaluation of sales methods and incentives, etc.
12. Qualitative research is an understanding of the perceptions, attitudes and predispositions of a certain target segment towards a product, service, advertising concept or a need area.
13. The information is used to identify and define marketing opportunities and problems; generate, refine and evaluate marketing actions; monitor marketing performance; and improve understanding of marketing as a process.
14. The following are the two courses which a company intending to buy marketing research may adopt. First, it may straightaway ask a marketing research agency to prepare a research proposal including an estimate of cost. Second, it may give some thought to its problem, spell it out and provide certain specifications according to which the concerned agency has to carry out the research.
15. Expenditure on unsuccessful product launches, wrong product positioning, improper distribution system etc. can whittle away the vitals of an organization.
16. Effective marketing starts with a strong knowledge of customers: the kind of knowledge that gives an organization unique insights into what customers want and how to satisfy them better than the competition.
17. Before the advent of internet technologies, a marketer had to make an important trade-off about the type of information he could send and the media by which he could send it.

1.9 SUMMARY

- Fred Kerlinger (1986) stated that ‘Scientific research is a systematic, controlled and critical investigation of propositions about various phenomena.’
- The world of business demands that managers and researchers work towards a goal—whether immediate or futuristic, else the research loses its significance in the field of management.
- Researchers use scientific methods and techniques to provide an optimum solution to the problems. The scientific methods include observing and defining a problem and formulating a hypothesis related to the results of the scientific methods and techniques.
- Exploratory research is used to gain a deeper understanding of the issue or problem that is troubling the decision maker. The idea is to provide direction to subsequent and more structured and rigorous research.

- Conclusive research is carried out to test and validate the study hypotheses. In contrast to exploratory research, these studies are more structured and definite. The variables and constructs in the research are clearly defined.
- The main goal of descriptive research is to describe the data and characteristics about what is being studied. The annual census carried out by the Government of India is an example of descriptive research.
- Causal research studies explore the effect of one thing on another and more specifically, the effect of one variable on another.
- Marketing is essentially a sum total of two entities, viz. the target market and the marketing mix for that market, which in turn is related to actions regarding the four Ps, viz., Product, Price, Place and Promotion.
- Marketing management functions, viz., planning, implementation, and control, relating to marketing mix are also performed in an effective manner only when information on elements of marketing-mix and environmental forces are made available to marketing executives operating at various levels of management.
- G Zaltman and P.C. Burger have criticized the AMA's definition of marketing research on the grounds that it does not include the pre-research analysis which is necessary to ascertain the type of information that should be collected, recorded and analyzed.
- Marketing research describes the situation, the background of any marketing problem. It helps us to understand in a preliminary manner the underlying marketing problems.
- Marketing research does not supplant or replace the decision-making process. It only serves as an input to the decision-making process.
- Market Research covers the different aspects, such as the size and nature of the market including export markets, dividing the consumers in terms of their age, gender and income (market segmentation).
- Problems relating to input and analysis, forecasting, price and profit analysis, cost-volume-profit analysis (break-even charts), etc., are the main fields of business economics research.
- Growth areas of marketing research include: (i) New product research, (ii) Financial research, (iii) Economic research, (iv) Service sector research, (v) Corporate image research, (vi) Industrial products research.
- Since 1992 liberal licensing and import policy of the Indian government has brought about many changes in the corporate environment.
- Effective marketing starts with a strong knowledge of customers: the kind of knowledge that gives an organization unique insights into what customers want and how to satisfy them better than the competition.
- Internet has changed the economics of information. Now it is possible to transfer brief as well as detailed messages to both small and wide audiences.

NOTES

NOTES

1.10 KEY WORDS

- **Fundamental Research:** It is a type of scientific research with the aim of improving scientific theories for better understanding and prediction of natural or other phenomena.
- **Applied Research:** This type of research uses scientific theories to develop technology or techniques which can be used to intervene and alter natural or other phenomena.
- **Exploratory Research:** It is the preliminary research to clarify the exact nature of the problem to be solved. It is used to ensure additional research is taken into consideration during an experiment as well as determining research priorities which may be difficult to take note of without exploratory research.
- **Marketing Information System (MIS):** It is a continuing and interacting structure of people, equipment and procedures to gather, sort, analyze, evaluate, and distribute pertinent, timely and accurate information for use by marketing decision makers to improve their marketing planning, implementation, and control.

1.11 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. State two characteristics of research.
2. List some examples of descriptive research.
3. What is situational analysis?
4. Write a short note on AMA's definition of marketing research.
5. Why has marketing research been in great demand with firms making/marketing consumer goods?
6. How is market research a narrow term when compared with marketing research?
7. How does economics of information put constraints on the development of the most efficient methods to deliver products and services?

Long-Answer Questions

1. Discuss conclusive research and its types.
2. Examine the nature of marketing research.
3. Analyse the objectives of marketing research.
4. Explain the importance of marketing research.

1.12 FURTHER READINGS

- Saravanel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.
- Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.
- Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.
- Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.
- Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

*Marketing Research
Dynamics*

NOTES

*Self-Instructional
Material*

NOTES

UNIT 2 PLANNING THE RESEARCH PROCESS

Structure

- 2.0 Introduction
- 2.1 Objectives
- 2.2 Overview of the Research Process
- 2.3 Components of the Research Proposal
- 2.4 Components of a Research Paper
- 2.5 Answers to Check Your Progress Questions
- 2.6 Summary
- 2.7 Key Words
- 2.8 Self Assessment Questions and Exercises
- 2.9 Further Readings

2.0 INTRODUCTION

The research process is carried out in a sequential order where a series of steps are performed. A sequential order helps in making the process well-structured and ordered thereby reducing the risk of errors. The process begins with the selection of the topic and is followed by steps such as defining the research process, mentioning the objectives of research, surveying existing literature, developing a working hypothesis, preparing the research design and so on. The main aim of research is to achieve the intended objectives by gathering information and details about the ways of its attainment. This unit provides an overview of the research process and the components of both research proposal and research paper.

2.1 OBJECTIVES

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

- Discuss the Research Process
- Analyse the components of a Research Proposal
- Describe the components of a Research Paper

2.2 OVERVIEW OF THE RESEARCH PROCESS

The process of research is implemented as a series of actions or steps that are essentially performed in a specific order. These actions or activities usually overlap each other rather than pursuing a specific sequence. A brief description of the steps is given as follows:

- **Selecting the topic:** The first step of a researcher is to select a topic of research. While doing so, he should restrict himself to the most potential topic that is open for extensive research out of several alternatives. The factors to be considered for topic selection are:

- o Relevance
- o Scope for research, i.e. the required data should be available and accessible
- o Contribution to knowledge in the specific field
- o Required cooperation from the research guide

- **Formulation of a problem:** There are two types of research problems:

- o Problems related to the state of nature
- o Problems related to the relationship of variables

In defining the research problem, the researcher should study the existing literature like books and journals, available in the field with an interdisciplinary perspective to base his research topic on some reliable background. He should also concentrate on the relevance of the present research with the past works.

- **Mention the objective of research:** After selecting the topic and defining the research problem, the researcher should mention the objective of research. This means that he should explain what he aims to achieve through the research. His objective should also explain the extent to which the research work is related to the specific field.

- **Survey existing literature:** To understand the basis of research, it is important for the researcher to review the existing literature. This involves:

- o Surveying the existing books available in the field
- o Reviewing other published literature like articles, journals, reports, conference proceedings etc.

The researcher should then prepare his own index for a period, in chronological order, in addition to his consultation of various indices.

- **Development of working hypothesis:** A hypothesis is an uncertain statement that involves the proposed answer to the problem. The hypothesis statement provides high priority to accountability and responsibility of research procedure. The solution proposed by working hypothesis cannot be considered as the only solution to the research problem. It only acts as the chain or interface between the theory and the research problem. It can also be considered as the point of departure. Hypotheses are thus the tentative statements that can either be rejected or accepted after the research process. Hypothesis also provides a structure and guides the researcher in the direction he should move to reach the solution of the problem. The

NOTES

NOTES

researcher must keep the following things in mind while formulating a working hypothesis:

- o Hypothesis can only be developed after the researcher is certain about the nature, extent and intensity of the problem.
 - o Hypothesis should be figured out throughout the research process which provides appropriate structure to the problem.
 - o The researcher should keep it in mind that hypothesis is only the tentative statements/solutions of the problem and this hypothesis should not be generalized much.
 - o A research problem does not need to have only one hypothesis. It only depends on the research proposal that how many hypotheses are required to solve the problem.
- **Preparing the research design:** Once the researcher has gained enough knowledge about the problem statement, he needs to prepare the plan that will act as the outline of the investigation in research process. The research design consists of a series of steps that has to be carried out during research.

There are two types of research design:

(i) Exploratory research

(ii) Conclusive research

(i) **Exploratory research:** The researcher conducts exploratory research when the problem has not been defined or he has not gained much knowledge about the research problem. Exploratory research allows the researcher to become familiar with the problem or the concept to be studied. The researcher can determine the best research design, data collection method and selection of subjects with this research. Sometimes, this research can also conclude that the problem does not exist. Exploratory research can be quite informal and can rely on secondary research and qualitative research.

(ii) **Conclusive research:** As specified by its name, it is used to provide information that can help the researcher in decision-making or in reaching a conclusion. This search is likely to be quantitative in nature. It depends on both secondary data, which is also called existing data, and primary research, or data that is collected for the current study only.

(a) *Descriptive research:* Descriptive research, also called statistical research, provides data about the population or universe that has to be studied during research. Descriptive research provides information about 'who, what, when, where and how' of a situation but it does not provide information about who caused the problem. The researcher can use the descriptive research when the objective is to provide systematic, accurate

and factual description. There are two types of descriptive research designs:

- Observations
- Surveys

(b) *Causal research*: It is used to find out the variable causing certain behaviour. This research is applicable when the researcher has the knowledge of variables that are causing the problem and that are affected by the problem. This type of research tends to be very complex and the researcher may sometimes be unable to determine the attitude of an individual by this research. There are two types of causal research designs:

- Experimentation
- Simulation

- **Determine the sample design**: Often only a few items are selected for universal study purposes, for example, blood testing on a sample basis to perform census inquiry. The items selected are technically known as a sample. The researcher must decide the way of selecting a sample or decide about sample design. A sample design is a definite plan determined for data collection to obtain a sample from a given population. The various types of sample designs are as follows:

- o Deliberate sampling
- o Simple random sampling
- o Systematic sampling
- o Stratified sampling
- o Quota sampling
- o Cluster sampling
- o Multi-stage sampling
- o Sequential sampling

The researcher should decide the sample design after considering the nature of inquiry and other related factors. Sometimes several above-mentioned methods of sampling are used in the same study, which is called mixed sampling.

- **Data collection methods**: There are a variety of ways to collect data. Primary data can be collected through experiments or through surveys. If the researcher performs an experiment, he observes some quantitative measurements. This helps him examine the validity of his hypothesis. In the case of surveys, however, the researcher can adopt one or more of the following ways to collect data:
 - o By observation
 - o Through personal interviews

NOTES

NOTES

- o Through telephone interviews
- o By mailing of questionnaires
- o Through schedules

- **Execution of project:** This is the most important step in the research process. The researcher should ensure that the project is performed in a logical way and in time. If a survey is to be carried out, steps should be taken to ensure that it is under statistical control, so that the collected data is in accordance with the pre-determined standard of accuracy.
- **Analysis of data:** After data collection, the researcher turns to the task of analyzing it. The bulk data should be compressed into a few manageable groups and tables for further analysis. The researcher can then analyze the collected data by using various statistical measures.
- **Hypothesis testing:** After analyzing the data, the researcher should test the hypothesis, if any. He should check if the facts support the hypothesis or are contrary. Statisticians have developed tests like Chi square test, t-test and F-test for hypothesis testing. This testing further results in either acceptance or rejection of hypothesis.
- **Generalizations and interpretations:** The real value of research lies in its ability to arrive at certain generalizations. If the researcher cannot find a hypothesis to start with, he might seek to explain his findings on the basis of some theory. This is called interpretation. This may give rise to new questions and lead to further research.
- **Report writing:** This is the concluding step of research, where the researcher has to prepare a report of what has been done by him. Generally, the report should be designed in accordance to the following layout:
 - o **The preliminary pages:** Here the title, date, acknowledgements and foreword with the table of contents, should be mentioned.
 - o **The main text:** This should be divided into introduction, summary, main report and conclusion.
 - o **End matter:** This should contain appendices, bibliography and index.

A report should be written in a precise and objective style in simple language. Charts and illustrations should be included to lay emphasis on the study of research.

Check Your Progress

1. What is the first step of a researcher?
2. When does the researcher conduct exploratory research?
3. When is causal research applicable?
4. What is interpretation?
5. What does the end matter of a report contain?

2.3 COMPONENTS OF THE RESEARCH PROPOSAL

Planning the Research
Process

We have learnt in this unit that research always begins with a purpose. Either this is the researcher's own pursuit, or it is carried out to address and answer a specific managerial question and arrive at an applicable solution. This clear statement of purpose guides the research process; however, as emphasized in the last section, for a study to qualify as research, it must be planned and systematic. Thus, the researcher needs to formalize this plan of pursuing the study. This framework or plan is termed as the research proposal. *A research proposal is a formal document that presents the research objectives, design of achieving these objectives and the expected outcomes/deliverables of the study.*

This step is essential both for academic and corporate research, as it clearly establishes the researcher's conceptualization of the research process that is intended to address the research questions. Through this written document the reader (academic expert or manager) is able to assess the rigour and validity of the study and whether or not it will result in an objective and accurate answer to the research problem. In a business or corporate setting, this step is often preceded by a PR (Proposal Request). Here the manager or the corporate spells out his decision problem and objectives and requests the potential suppliers of research to work out a research plan/proposal to address the stated issues. Thus, the research proposal submitted in such cases allows the manager to assess the credentials of the research agency or researcher as well as the proposed plan and to compare them with other proposals submitted. Then the manager selects the one that he feels would be able to most effectively (in terms of cost, time and accuracy) achieve the stated research goals.

Another advantage of a formal proposal is that sometimes the manager may not be able to clearly identify or enunciate his problem or the researcher might not be able to comprehend and convert the decision into a viable and workable research problem. The researcher lists out the objectives of the study and then together with the manager, is able to review whether or not the listed objectives and direction of the study will be able to deliver the necessary inputs required for arriving at a workable solution.

For the researcher, the document provides an opportunity to identify any shortfalls in the logic or the assumption of the study. When the researcher defines the flow and order of the steps required in the research process, he is also creating a mechanism for identifying probabilities of possible interrelated or simultaneous activities that can be carried out. It also helps to monitor the methodical work being carried out to accomplish the project.

Basically the proposals formulated could be of three types. The first is the academic research proposal that might be generated by students or academicians pursuing the study for fundamental academic research. An example is an

NOTES

NOTES

academician wanting to explore the viability of different eco-friendly packaging options available to a manufacturer.

The second type of proposals are internal to an organization and are submitted to the management for approval and funding. They are of a highly focused nature and are oriented towards solving immediate problems. For example, a pharmaceutical company, which has developed a new hair growing formulation; wants to test whether to package the liquid in a spray type or capped dispenser. The solutions are time-driven and applicability is only for this product. These studies do not require extensive literature review but do require clearly stated research objectives, for the management to assess the nature of work required.

The third type of proposals have the base or origin within the company, but the scope and nature of the study requires a more structured and objective research. For example, if the above stated pharmaceutical company wishes to explore the herbal cosmetic market and wants market analysis and feasibility study conducted; the PR might be spelt out to solicit proposals to address the research question, and execute an outsourced research.

Contents of a Research Proposal

As stated above, the requirements and the origin of the research would direct the sequential formulation of the research proposal. However, there is a broad framework that most proposals adhere to. In this section we will briefly discuss these steps.

(i) Executive Summary

This is a broad overview or abstract that spells out the purpose and objective of the study. In a short paragraph, the author gives a summary about the management problem/academic concern, which is the backdrop of the study. The probable research questions which might need to be answered in order to arrive at any conclusive results are further listed.

(ii) Background of the Problem

This is the detailed background of the management problem. It requires a sequential and systematic build-up to the research questions and also a compelling reason for pursuing the study. The researcher has to be able to demonstrate that there could be a number of ways in which the management dilemma could be addressed. For example, in the pharmaceutical company, the product testing could be done internally in the company, or the two sample bottles could be formulated and tested for their acceptability amongst probable consumers or retailers stocking the product; or the two prototypes would be developed and test launched and tested for their sales potential. The researcher thus has to spell out all probabilities and then systematically and logically argue for the intended research study. This section has to be explicit, objective and written in simple language, avoiding any metaphors or idioms to dramatize the plan. The logical arguments should speak for themselves

and be able to convince the reader of the need for the study in order to find probable solutions to the management dilemma.

(iii) Problem Statement and Research Objectives

The clear definition of the problem broken down into specific objectives is the next step. This section is crisp and to the point. It begins by stating the main thrust area of the study. For example, in the above case, the problem statement could be:

To test the acceptability of a spray or capped bottle dispenser for a new hair growing formulation. The basic objectives of this research would be to:

- Determine the comparative preference of the two prototypes amongst customers of hair growing solutions
- To conduct a sample usage test of both the bottles with the identified population
- To assess the ease of use for the bottles amongst the respondents
- To prepare a comparative analysis of the advantages and problems associated with each bottle, on the basis of the sample usage test
- To prepare a detailed feasibility report on the basis of the findings

If the study is addressed towards testing some assumptions in the form of hypotheses, they have to be clearly stated in this section.

(iv) Research Design

This is the working section of the proposal as it needs to indicate the logical and systematic approach intended to be followed in order to achieve the listed objectives. This would include specifying the population to be studied, the sampling process and plan, sample size and selection. It also details the information areas of the study and the probable sources of data, i.e., the data collection methods. In case the process has to include an instrument design, then the intended approach needs to be detailed here. A note of caution has to be given here, this is not a simple statement of the sampling and data collection plan, it requires a clear and logical justification of using the techniques over a wide gamut of methods available for research. For example, in the pharmaceutical study—a before and after design, a respondent population of customers who use like products and the use of a structured questionnaire over other methods, have to be justified.

(v) Scheduling the Research

The time-bound dissemination of the study with the major phases of the research has to be presented. This can be done using the CPM/GANTT/PERT charts. This gives a clear mechanism for monitoring and managing the research task. It also has the additional benefit of providing the researcher with a means of spelling out the payment points linked to the delivered phase outputs.

NOTES

NOTES

(vi) Results and Outcomes of the Research

Here the clear terms of contract or expected outcomes of the study have to be spelt out. This is essential even if it is an academic research. The expected deliverables need to clearly demonstrate how the researcher intends to link the findings of the proposed study design to the stated research objectives. For example, in the pharmaceutical study, the expected deliverables are:

- To identify the usage problems with each bottle type.
- To recommend on the basis of the sample study which bottle to use for packaging the liquid.

(vii) Costing and Budgeting the Research

In all instances of business research, both internal and external, an estimated cost of the study is required. A typical sample budget format with payment schedules is presented in the following sample proposal.

In addition to these sections, academic research requires a review of related literature section; this generally follows the ‘problem background’ section. If the proposal is meant to establish the credentials of the research supplier, then detailed qualifications of the research team, including the research experience in the required or related area, help to aid in the selection of the research proposal.

Sometimes, the research study requires an understanding of some technical terms or explanations of the constructs under study; in such cases, the researcher needs to attach a glossary of terms in the appendix of the research proposal.

The last section of the proposal is to state the complete details of the references used in the formulation of the research proposal. Thus the data source and address has to be attached with the formulated document.

Check Your Progress

6. State one advantage of a formal proposal.
7. What is an executive summary?
8. What are the benefits of scheduling a research?

2.4 COMPONENTS OF A RESEARCH PAPER

Research paper is nothing but the report that is presented at the end of the research. Before you write a report, you should define the high level structure of the report. Defining a clear logical structure will make the report easier to write and to read. There are two types of report structures, which are listed as follows:

- **Report structure I:** In general, the report writing structure comprises the following subheadings:
 - o Title Page
 - o Abstract

- o Table of Contents
- o Introduction
- o Technical Detail and Results
- o Discussion and Conclusions
- o References
- o Appendices
- **Report structure II:** There is also a specific structure of report writing pertaining to technical or scientific reports which is as follows:
 - o Introduction
 - o Background and Context
 - o Technical Details
 - o Results
 - o Discussion and Conclusion
- **Order of writing:**
 - o Start with the technical chapters/sections.
 - o Follow with the discussion.
 - o Finally, write the conclusions, introduction and abstract, if you are including any.
- **Appendix:** The appendix should contain the following:
 - o Material that suits or goes well with the flow of the main report but cannot be included in the main text of the report either because it is too long or is not an essential reading, for example, lists of parameter values, etc.
 - o Bibliography, i.e., list of all the sources of material, you referred to in your report.

You will learn more about report writing in Unit 12.

Check Your Progress

9. List the subheadings of a report writing structure.
10. What should an appendix contain?

2.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The first step of a researcher is to select a topic of research. While doing so, he should restrict himself to the most potential topic that is open for extensive research out of several alternatives.

NOTES

NOTES

2. The researcher conducts exploratory research when the problem has not been defined or he has not gained much knowledge about the research problem.
3. Causal research is applicable when the researcher has the knowledge of variables that are causing the problem and that are affected by the problem.
4. If the researcher cannot find a hypothesis to start with, he might seek to explain his findings on the basis of some theory. This is called interpretation.
5. The end matter of a research contains appendices, bibliography and index.
6. The advantage of a formal proposal is that the researcher lists out the objectives of the study and then together with the manager, is able to review whether or not the listed objectives and direction of the study will be able to deliver the necessary inputs required for arriving at a workable solution.
7. Executive summary is a broad overview or abstract that spells out the purpose and objective of the study. In a short paragraph, the author gives a summary about the management problem/academic concern, which is the backdrop of the study.
8. Scheduling a research gives a clear mechanism for monitoring and managing the research task. It also has the additional benefit of providing the researcher with a means of spelling out the payment points linked to the delivered phase outputs.
9. The subheadings of a report writing structure are: (i) Title Page, (ii) Abstract, (iii) Table of Contents, (iv) Introduction, (v) Technical Detail and Results, (vi) Discussion and Conclusions, (vii) References, (viii) Appendices
10. An appendix should contain the following:
 - Material that suits or goes well with the flow of the main report but cannot be included in the main text of the report either because it is too long or is not essential reading, for example, lists of parameter values, etc.
 - Bibliography, i.e., list of all the sources of material, you referred to in your report.

2.6 SUMMARY

- The process of research is implemented as a series of actions or steps that are essentially performed in a specific order. These actions or activities usually overlap each other rather than pursuing a specific sequence.
- After selecting the topic and defining the research problem, the researcher should mention the objective of research. This means that he should explain what he aims to achieve through the research.

- Exploratory research allows the researcher to become familiar with the problem or the concept to be studied. The researcher can determine the best research design, data collection method and selection of subjects with this research.
- Descriptive research, also called statistical research, provides data about the population or universe that has to be studied during research.
- The researcher should decide the sample design after considering the nature of inquiry and other related factors.
- After analyzing the data, the researcher should test the hypothesis, if any. He should check if the facts support the hypothesis or are contrary.
- The specific way in which a variable is measured in a particular study is called the operational definition. It is critical to operationally define a variable in order to lend credibility to the methodology and to ensure the reproducibility of the results.
- Through the research proposal, a reader (academic expert or manager) is able to assess the rigour and validity of the study and whether or not it will result in an objective and accurate answer to the research problem.
- When the researcher defines the flow and order of the steps required in the research process, he is also creating a mechanism for identifying probabilities of possible interrelated or simultaneous activities that can be carried out.
- The researcher has to spell out all probabilities and then systematically and logically argue for the intended research study.
- Research design is the working section of the proposal as it needs to indicate the logical and systematic approach intended to be followed in order to achieve the listed objectives. This would include specifying the population to be studied, the sampling process and plan, sample size and selection.
- The time-bound dissemination of the study with the major phases of the research has to be presented. This can be done using the CPM/GANTT/PERT charts.
- In all instances of business research, both internal and external, an estimated cost of the study is required. A typical sample budget format with payment schedules is presented in the following sample proposal.
- Sometimes, the research study requires an understanding of some technical terms or explanations of the constructs under study; in such cases the researcher needs to attach a glossary of terms in the appendix of the research proposal.

NOTES

2.7 KEY WORDS

- **Hypothesis:** It refers to a supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation.

NOTES

- **Operational Definition:** It is the specific way in which a variable is measured in a particular study.
- **Sample Design:** It is a definite plan determined for data collection to obtain a sample from a given population.
- **Research Proposal:** It is a formal document that presents the research objectives, design of achieving these objectives and the expected outcomes/deliverables of the study.

2.8 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What should a researcher study for defining a research problem?
2. Write a short note on descriptive research.
3. List the various types of sample design.
4. What should be the layout of a report?
5. Write a short note on research design.

Long-Answer Questions

1. Discuss the steps involved in the research process.
2. Explain the three types of formal proposals.

2.9 FURTHER READINGS

- Saravanel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.
- Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.
- Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.
- Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.
- Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

UNIT 3 RESEARCH DESIGN

Structure

- 3.0 Introduction
- 3.1 Objectives
- 3.2 Meaning of Research Design
- 3.3 Types of Research Design: Descriptive Research and Causal Research Design
- 3.4 Research Design and Marketing Decision Process
- 3.5 Choosing a Good Research Design
 - 3.5.1 Selection of Research Design
- 3.6 Answers to Check Your Progress Questions
- 3.7 Summary
- 3.8 Key Words
- 3.9 Self Assessment Questions and Exercises
- 3.10 Further Readings

NOTES

3.0 INTRODUCTION

Research design refers to a strategy or method used to acquire the required information. It is a proper framework of procedures to be followed for collecting the information by deciding which sources are to be used. It is essential that the strategy is relevant to the research question. The important characteristics of research design are reliability, neutrality, validity and generalization. Descriptive research design and experimental design are two important research designs that are used for attaining the information needed for significant objectives. The objective of a descriptive research study is to provide a comprehensive and detailed explanation of the phenomena under study whereas experimental designs are conducted to infer causality. This unit provides an analysis of research design, its types, marketing decision process and the components of choosing a good research design.

3.1 OBJECTIVES

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

- Describe the meaning of Research Design
- Discuss the types of Research Design
- Analyse Research Design and Marketing Decision Process
- Describe the components of a good Research Design

3.2 MEANING OF RESEARCH DESIGN

Once you have established the ‘what’ of the study, i.e., the research problem, the next step is the ‘how’ of the study, which specifies the method of achieving the research objectives. In other words, this is the research design.

NOTES

Green *et al.* (2008) defines research design as ‘the specification of methods and procedures for acquiring the information needed. It is the overall operational pattern or framework of the project that stipulates what information is to be collected from which sources by what procedures. If it is a good design, it will ensure that the information obtained is relevant to the research questions and that it was collected by objective and economical procedures.’

Thyer (1993) states that, ‘A traditional research design is a blueprint or detailed plan for how a research study is to be completed—operationalizing variables so they can be measured, selecting a sample of interest to study, collecting data to be used as a basis for testing hypotheses, and analyzing the results.’ Sellitz *et al.* (1962) state that, ‘A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure.’

One of the most comprehensive and holistic definition has been given by Kerlinger (1995). He refers to a research design as, ‘..... a plan, structure and strategy of investigation so conceived as to obtain answers to research questions or problems. The plan is the complete scheme or programme of the research. It includes an outline of what the investigator will do from writing the hypotheses and their operational implications to the final analysis of data.’

Thus, the formulated design must ensure three basic principles:

- (a) Convert the research question and the stated assumptions/hypotheses into variables that can be measured.
- (b) Specify the process to complete the above task.
- (c) Specify the ‘control mechanism(s)’ to follow so that the effect of other variables that could have an effect on the outcome of the study have been controlled.

At this stage, one needs to understand the difference between research design and research method. While the design is the specific framework that has been created to seek answers to the research question, the research method is the technique to collect the information required to answer the research problem, given the created framework. Thus, research designs have a critical and directive role to play in the research process. The execution details of the research question to be investigated are referred to as the research design.

The researcher has a number of designs available to him for investigating the research objectives. The classification that is universally followed is the one based upon the objective or the purpose of the study. A simple classification that is based upon the research needs ranging from simple and loosely structured to the specific and more formally structured. The best way is to view the designs on a continuum as shown in Figure 3.1. Hence, in case the research objective is diffused and requires a refinement, one uses the exploratory design, and this might lead to the slightly more concrete descriptive design—here one describes all the aspects of the construct and concepts under study. This leads to a more structured and controlled experimental research design.

Figure 3.1 illustrates research designs as a continuous process.

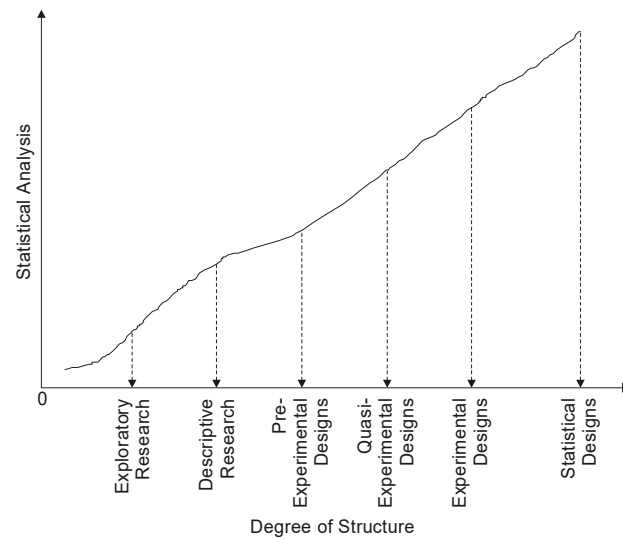


Fig. 3.1 Research Designs—A Continuous Process

In the following sections, you will study the broad classification of research design.

Check Your Progress

1. How does Kerlinger define research design?
2. How do research design and research method differ?

3.3 TYPES OF RESEARCH DESIGN: DESCRIPTIVE RESEARCH AND CAUSAL RESEARCH DESIGN

Research designs as mentioned in Unit 2 are broadly divided into exploratory and conclusive research. Descriptive and causal research are a part of conclusive research. In this section, let's first briefly study exploratory research designs and then learn about conclusive research designs.

I. Exploratory Designs

Exploratory designs, as stated earlier, are the simplest and most loosely structured designs. As the name suggests, the basic objective of the study is to explore and obtain clarity about the problem situation. It is flexible in its approach and mostly involves a qualitative investigation. The sample size is not strictly representative and at times it might only involve unstructured interviews with a couple of subject experts. The essential purpose of the study is to:

- Define and understand the research problem to be investigated.
- Explore and evaluate the diverse and multiple research opportunities.

NOTES

NOTES

- Assist in the development and formulation of the research hypotheses.
- Define the variables and constructs under study.
- Identify the possible nature of relationships that might exist between the variables under study.
- Explore the external factors and variables that might impact the research.

For example, a university professor might decide to do an exploratory analysis of the new channels of distribution that are being used by the marketers to promote and sell products and services. To do this, a structured and defined methodology might not be essential as the basic objective is to understand how to teach this to students of marketing. The researcher can make use of different methods and techniques in an exploratory research- like secondary data sources, unstructured or structured observations, expert interviews and focus group discussions with the concerned respondent group. Here, we will discuss them in brief in the light of their use in exploratory research.

(i) Secondary Resource Analysis

Secondary sources of data, as the name suggests, are data in terms of the details of previously collected findings in facts and figures—which have been authenticated and published. It is a fast and inexpensive way of collecting information. The past details can sometimes point out to the researcher that his proposed research is redundant and has already been established earlier. Secondly, the researcher might find that a small but significant aspect of the concept has not been addressed and should be studied. For example, a marketer might have extensively studied the potential of the different channels of communication for promoting a ‘home maintenance service’ in Greater Mumbai. However, there is no impact of any mix that he has tested. An anthropologist research associate, on going through the findings, postulated the need for studying the potential of WOM (word of mouth) in a close-knit and predominantly Parsi colony where this might be the most effective culture-dependent technique that would work. Thus, such insights might provide leads for carrying out an experimental and conclusive research subsequently.

Another valuable secondary resource is the compiled and readily available databases of the entire industry, business or construct. These might be available on free and public domains or through a structured acquisition process and cost. These are both government and non-government publications. Based on the resources and the level of accuracy required, the researcher might decide to make use of them.

(ii) Case Study Method

Another way of conducting an exploratory research is the case study method. This requires an in-depth study and is focused on a single unit of analysis. This unit could be an employee or a customer; an organization or a complete country analysis.

They are by their nature, generally, post-hoc studies and report those incidences which might have occurred earlier. The scenario is reproduced based upon the secondary information and a primary interview/discussion with those involved in the occurrence. Thus, there might be an element of bias as the data, in most cases, becomes a judgemental analysis rather than a simple recounting of events.

For example, BCA Corporation wants to implement a performance appraisal system in the organization and is debating between the merits of a traditional appraisal system and a 360° appraisal system. For a historical understanding of the two techniques, the HR director makes use of books on the subject. However, for better understanding, he should do an in-depth case accounting of Allied Association which had implemented traditional appraisal formats, and Surakhsha International which uses 360° appraisal systems. Thus, the two exploratory researches carried out were sufficient to arrive at a decision in terms of what would be best for the organization.

(iii) Expert Opinion Survey

At times, there might be a situation when the topic of a research is such that there is no previous information available on it. In these cases, it is advisable to seek help from experts who might be able to provide some valuable insights based upon their experience in the field or with the concept. This approach of collecting particulars from significant and knowledgeable people is referred to as the expert opinion survey. This methodology might be formal and structured and is useful when authenticated or supported by a secondary/primary research or it might be fluid and unstructured and might require an in-depth interviewing of the expert. For example, the evaluation of the merit of marketing organic food products in the domestic Indian market cannot be done with the help of secondary data as no such structured data sources exist. In this case the following can be contacted:

- Doctors and dieticians as experts would be able to provide information whether consumers would eat organic food products as a healthier alternative.
- Chefs who are experimental and would like to look at providing better value to their clients.
- Retailers who like to sell contemporary new products.

These could be useful in measuring the viability of the proposed plan. Discussions with knowledgeable people may reveal some information regarding who might be considered as potential consumers. Secondly, the question whether a healthy proposition or a lifestyle proposition would work better to capture the targeted consumers needs to be examined. Thus, this method can play a directional role in shaping the research study.

(iv) Focus Group Discussions

Another way to conduct an exploratory analysis is carry out discussions with individuals associated with the problem under study. This technique, though

NOTES

NOTES

originally from sociology, is actively used in business research. In a typical focus group, there is a carefully selected small set of individuals representative of the larger respondent population under study. It is called a focus group as the selected members discuss the concerned topic for the duration of 90 minutes to, sometimes, two hours. Usually the group is made up of six to ten individuals. The number thus stated is because less than six would not be able to throw enough perspectives for the discussion and there might emerge a one-sided discussion on the topic. On the other hand, more than ten might lead to more confusion rather than any fruitful discussion and that would be unwieldy to manage. Generally, these discussions are carried out in neutral settings by a trained observer, also referred to as the moderator. The moderator, in most cases, does not participate in the discussion. His prime objective is to manage a relatively non-structured and informal discussion. He initiates the process and then maneuvers it to steer it only to the desired information needs. Sometimes, there is more than one observer to record the verbal and non-verbal content of the discussion. The conduction and recording of the dialogue requires considerable skill and behavioural understanding and the management of group dynamics. In the organic food product study, the focus group discussions were carried out with the typical consumers/buyers of grocery products. The objective was to establish the level of awareness about health hazards, environmental concerns and awareness of organic food products. A series of such focus group discussions carried out across four metros—Delhi, Mumbai, Bengaluru and Hyderabad—revealed that even though the new age consumer was concerned about health, the awareness about organic products varied from extremely low to non-existent. *(This study was carried out in the year 2004–05 by one of the authors for an NGO located in Delhi).*

II Conclusive Research

The findings and propositions developed as a consequence of exploratory research might be tested and authenticated by conclusive research. This kind of research study is especially carried out to test and validate formulated hypotheses and specified relationships. In contrast, to exploratory research, these studies are more structured and definite. The variables and constructs in the research are clearly defined with explicit quantifiable indications or simply, the variables can be denoted in the form of numbers that can be quantified and summarized. As mentioned earlier, this type of research design is sub-divided into descriptive and causal research designs.

(i) Descriptive Research Designs

As the name implies, the objective of descriptive research studies is to provide a comprehensive and detailed explanation of the phenomena under study. The intended objective might be to give a detailed sketch or profile of the respondent population being studied. For example, to design an advertising and sales promotion campaign for high-end watches, a marketer would require a holistic profile of the

population that buys such luxury products. Thus a descriptive study, (which generates data on *who*, *what*, *when*, *where*, *why* and *how* of luxury accessory brand purchase) would be the design necessary to fulfill the research objectives.

Descriptive research thus are conclusive studies. However, they lack the precision and accuracy of experimental designs, yet it lends itself to a wide range of situations and is more frequently used in business research. Based on the time period of the collection of the research information, descriptive research is further subdivided into two categories: cross-sectional studies and longitudinal studies.

NOTES

(a) Cross-sectional Studies

As the name suggests, cross-sectional studies involve a slice of the population. Just as in scientific experiments one takes a cross-section of the leaf or the cheek cells to study the cell structure under the microscope, similarly one takes a current subdivision of the population and studies the nature of the relevant variables being investigated.

There are two essential characteristics of cross-sectional studies:

- The cross-sectional study is carried out at a single moment in time and thus the applicability is most relevant for a specific period. For example, one cross-sectional study was conducted in 2002 to study the attitude of Americans towards Asian-Americans, after the 9/11 terrorist attack. This revealed the mistrust towards Asians. Another cross-sectional study conducted in 2012 to study the attitude of Americans towards Asian-Americans revealed more acceptance and less mistrust. Thus the cross-sectional studies cannot be used interchangeably. .
- Secondly, these studies are carried out on a section of respondents from the population units under study (e.g., organizational employees, voters, consumers, industry sectors). This sample is under consideration and under investigation only for the time coordinate of the study.

There are also situations in which the population being studied is not of a homogeneous nature but composed of different groups. Thus it becomes essential to study the sub-segments independently. This variation of the design is termed as *multiple cross-sectional studies*. Usually this multi-sample analysis is carried out at the same moment in time. However, there might be instances when the data is obtained from different samples at different time intervals and then they are compared. *Cohort analysis* is the name given to such cross-sectional surveys conducted on different sample groups at different time intervals. Cohorts are essentially groups of people who share a time zone or have experienced an event that took place at a particular time period. For example, in the post-9/11 cross-sectional study done in 2002, we study and compare the attitudes of middle-aged Americans versus teenaged Americans towards Asian-Americans. These two American groups are separate cohorts and this would be a cohort analysis. Thus the teenage American is one cohort and the middle-aged cohort is separate and thinks differently.

NOTES

The technique is especially useful in predicting election results, cohorts of males–females, different religious sects, urban–rural or region-wise cohorts are studied by leading opinion poll experts like Nielsen, Gallup and others. Thus, Cross-sectionals studies are extremely useful to study current patterns of behaviour or opinion.

(b) Longitudinal Studies

A single sample of the identified population that is studied over a longer period of time is termed as a longitudinal study design. A panel of consumers specifically chosen to study their grocery purchase pattern is an example of a longitudinal design. There are certain distinguishing features of the same:

- The study involves the selection of a representative panel, or a group of individuals that typically represent the population under study.
- The second feature involves the repeated measurement of the group over fixed intervals of time. This measurement is specifically made for the variables under study.
- A distinguishing and mandatory feature of the design is that once the sample is selected, it needs to stay constant over the period of the study. That means the number of panel members has to be the same. Thus, in case a panel member due to some reason leaves the panel, it is critical to replace him/her with a representative member from the population under study.

Longitudinal study using the same section of respondents thus provides more accurate data than one using a series of different samples. These kinds of panels are defined as true panels and the ones using a different group every time are called omnibus panels. The advantages of a true panel are that it has a more committed sample group that is likely to tolerate extended or long data collecting sessions. Secondly, the profile information is a one-time task and need not be collected every time. Thus, a useful respondent time can be spent on collecting some research-specific information.

However, the problem is getting a committed group of people for the entire study period. Secondly, there is an element of mortality and attrition where the members of the panel might leave midway and the replaced new recruits might be vastly different and could skew the results in an absolutely different direction. A third disadvantage is the highly structured study situation which might be responsible for a consistent and structured behaviour, which might not be the case in the real or field conditions.

(ii) Causal Research

To address the need for establishing causality, there is another kind of conclusive research study called causal research. These studies establish the why and the how of a phenomenon. Causal research explores the effect of one thing on

another and more specifically, the effect of one variable on another. They are highly structured and require a rigid sequential approach to sampling, data collection and data analysis. The design of the study takes on a critical significance here. To establish a reliable and testable relationship between two or more constructs or variables, the other influencing variables must be controlled so that their impact on the effect can be eliminated or minimized. For example, to study the impact of flexible work policies on turnover intentions, the other intervening variables, of age, marital status, organizational commitment and job autonomy would need to be controlled. This method of controlling the intervening variables will be discussed in detail in the subsequent chapter. This kind of research, like research in pure sciences, requires experimentation to establish causality. In majority of the situations, it is quantitative in nature and requires statistical testing of the information collected.

Experimental Designs

Experimental designs are conducted to infer causality. In an experiment, a researcher actively manipulates one or more causal variables and measures their effects on the dependent variables of interest. Since any changes in the dependent variable may be caused by a number of other variables, the relationship between cause and effect often tends to be probabilistic in nature. It is virtually impossible to prove a causality. One can only infer a cause-and-effect relationship.

The necessary conditions for making causal inferences are: (i) concomitant variation, (ii) time order of occurrence of variables and (iii) absence of other possible causal factors. The first condition implies that cause and effect variables should have a high correlation. The second condition means that causal variable must occur prior to or simultaneously with the effect variable. The third condition means that all other variable except the one whose influence we are trying to study should be absent or kept constant.

There are two conditions that should be satisfied while conducting an experiment. These are:

- **Internal validity:** Internal validity tries to examine whether the observed effect on a dependent variable is actually caused by the treatments (independent variables) in question. For an experiment to be possessing internal validity, all the other causal factors except the one whose influence is being examined should be absent. Control of extraneous variables is a necessary condition for inferring causality. Without internal validity, the experiment gets confounded.
- **External validity:** External validity refers to the generalization of the results of an experiment. The concern is whether the result of an experiment can be generalized beyond the experimental situations. If it is possible to generalize the results, then to what population, settings, times, independent variables and the dependent variables can the results be projected. It is desired to have an experiment that is valid both internally and externally. However, in

NOTES

NOTES

reality, a researcher might have to make a trade-off between one type of validity for another. To remove the influence of an extraneous variable, a researcher may set up an experiment with artificial setting, thereby increasing its internal validity. However, in the process the external validity will be reduced.

There are four types of experimental designs. These are explained below:

- (a) **Pre-experimental designs:** There are three designs under this – one short case study where observation is taken after the application of treatment, one group pre test-post test design where one observation is taken prior to the application of treatment and the other one after the application of treatment, and static group comparison, where there are two groups – experimental group and control group. The experiment group is subjected to treatment and a post test measurement is taken. In the control group measurement is taken at the time when it was done for experimental group. These do not make use of any randomization procedures to control the extraneous variables. Therefore, the internal validity of such designs is questionable.
- (b) **Quasi-experimental designs:** In these designs, the researcher can control when measurements are taken and on whom they are taken. However, this design lacks complete control of scheduling of treatment and also lacks the ability to randomize test units' exposure to treatments. As the experimental control is lacking, the possibility of getting confounded results is very high. Therefore, the researchers should be aware of what variables are not controlled and the effects of such variables should be incorporated into the findings.
- (c) **True experimental designs:** In these designs, researchers can randomly assign test units and treatments to an experimental group. Here, the researcher is able to eliminate the effect of extraneous variables from both the experimental and control group. Randomization procedure allows the researcher the use of statistical techniques for analyzing the experimental results.
- (d) **Statistical designs:** These designs allow for statistical control and analysis of external variables. The main advantages of statistical design are the following:
 - The effect of more than one level of independent variable on the dependent variable can be manipulated.
 - The effect of more than one independent variable can be examined.
 - The effect of specific extraneous variable can be controlled.

Statistical design includes the following designs:

- *Completely randomized design:* This design is used when a researcher is investigating the effect of one independent variable on the dependent variable. The independent variable is required to be measured in nominal scale i.e. it should have a number of categories.

Each of the categories of the independent variable is considered as the treatment. The basic assumption of this design is that there are no differences in the test units. All the test units are treated alike and randomly assigned to the test groups. This means that there are no extraneous variables that could influence the outcome.

Suppose we know that the sales of a product is influenced by the price level. In this case, sales are a dependent variable and the price is the independent variable. Let there be three levels of price, namely, low, medium and high. We wish to determine the most effective price level i.e. at which price level the sale is highest. Here, the test units are the stores which are randomly assigned to the three treatment level. The average sales for each price level is computed and examined to see whether there is any significant difference in the sale at various price levels. The statistical technique to test for such a difference is called analysis of variance (ANOVA).

The main limitation of completely randomized designs is that it does not take into account the effect of extraneous variables on the dependent variable. The possible extraneous variables in the present example could be the size of the store, the competitor's price and price of the substitute product in question. This design assumes that all the extraneous factors have the same influence on all the test units which may not be true in reality. This design is very simple and inexpensive to conduct.

- *Randomized block design:* As discussed, the main limitation of the completely randomized design is that all extraneous variables were assumed to be constant over all the treatment groups. This may not be true. There may be extraneous variables influencing the dependent variable. In the randomized block design it is possible to separate the influence of one extraneous variable on a particular dependent variable, thereby providing a clear picture of the impact of treatment on test units.

In the example considered in the completely randomized design, the price level (low, medium and high) was considered as an independent variable and all the test units (stores) were assumed to be more or less equal. However, all stores may not be of the same size and, therefore, can be classified as small, medium and large size stores. In this design, the extraneous variable, like the size of the store could be treated as different blocks. Now the treatments are randomly assigned to the blocks in such a way so that each treatment appears in each block at least once. The purpose of forming these blocks is that it is hoped that the scores of the test units within each block would be more or less homogeneous when the treatment is absent. What is assumed here is that block (size of the store) is correlated with the dependent variable

NOTES

NOTES

(sales). It may be noted that blocking is done prior to the application of the treatment.

In this experiment one might randomly assign 12 small-sized stores to three price levels in such a way that there are four stores for each of the three price levels. Similarly, 12 medium-sized stores and 12 large-sized stores may be randomly assigned to three price levels. Now the technique of analysis of variance could be employed to analyse the effect of treatment on the dependent variable and to separate out the influence of extraneous variable (size of store) from the experiment.

- *Factorial design:* A factorial design may be employed to measure the effect of two or more independent variables at various levels. The factorial designs allow for interaction between the variables. An interaction is said to take place when the simultaneous effect of two or more variables is different from the sum of their individual effects. An individual may have a high preference for mangoes and may also like ice-cream, which does not mean that he would like mango ice cream, leading to an interaction.

The sales of a product may be influenced by two factors, namely, price level and store size. There may be three levels of price—low (A_1), medium (A_2) and high (A_3). The store size could be categorized into small (B_1) and big (B_2). This could be conceptualized as a two-factor design with information reported in the form of a table. In the table, each level of one factor may be presented as a row and each level of another variable would be presented as a column. This example could be summarized in the form of a table having three rows and two columns. This would require $3 \times 2 = 6$ cells. Therefore, six different levels of treatment combinations would be produced each with a specific level of price and store size. The respondents would be randomly selected and randomly assigned to the six cells.

The tabular presentation of 3×2 factorial design is given in Table 3.2.

Table 3.2 3×2 Factorial Design for Price Level and Store Size

Price	Store	
	Small (B_1)	Big (B_2)
Low Level (A_1)	A_1B_1	A_1B_2
Medium Level (A_2)	A_2B_1	A_2B_2
High Level (A_3)	A_3B_1	A_3B_2

Respondents in each cell receive a specified treatment combination. For example, respondents in the upper left hand corner cell would face small level of price and small store. Similarly, the respondents in

the lower right hand corner cell will be subjected to both high price level and big store.

The main advantages of factorial design are:

- It is possible to measure the main effects and interaction effect of two or more independent variables at various levels.
- It allows a saving of time and effort because all observations are employed to study the effects of each factor.
- The conclusion reached using factorial design has broader applications as each factor is studied with different combinations of other factors.

The limitation of this design is that the number of combinations (number of cells) increases with increased number of factors and levels. However, a fractional factorial design could be used if interest is in studying only a few of the interactions or main effects.

NOTES

Check Your Progress

3. Why is there an element of bias in the case study method?
4. What is an expert opinion survey?
5. What is the objective of descriptive research studies?
6. What are the necessary conditions for making causal inferences?
7. List two main advantages of factorial design.

3.4 RESEARCH DESIGN AND MARKETING DECISION PROCESS

The very essence of managerial action in any modern organisation is decision making. Such managerial decision making is commonly referred to as the formulation, evaluation and selection of alternatives to solve managerial problems. The primary activity of marketing management is making such decisions with regard to marketing problem. This chapter provides the framework for examining role of research in marketing decision making.

Levels of Decision Making

First, the levels of decision making in an organisation are identified and a generalised model of the decision making process is presented. Then its role in marketing management is examined. The diversity of decision making activities in marketing organisations is boundless. This diversity of decision making results in a variety of informational needs within an organisation. The types of informational needs are related to levels of decision making. The three major levels of decision making in the organisation are strategic, tactical and operational levels. *Strategic decisions*

NOTES

are those that are largely unstructured and which affect the general direction of the organisation. Examples are such issues as diversification, product or market development and divestiture decisions. Examples of *tactical decisions*, include budget allocations, promotional mix decisions and personnel assignments. *Operational decisions* are of a routine nature and deal with the control of specific tasks. Examples are quality control, scheduling and transportation.

Decision Making Process

The decision making process can be thought of as a series of interrelated activities that leads to choice among alternatives. These activities are defined and interrelated as follows:

- **Problem Recognition :** The acknowledgment by management that some situation exists or will exist, that need to be acted upon in the near future.
- **Information Search:** Information pertaining to the problem situation identified above is collected and assimilated into a format that is conducive to analysis of the problem at hand. Assumption and alternative scenarios are posed for subsequent problem analysis. The information generated at this stage may substantially redefine the problem initially formulated.
- **Problem Analysis:** The problem situation is thoroughly reviewed given the information obtained. An analysis of the available information reveals the areas of concern to the manager and the major factors that affect the problem at hand. This may also affect recognition and necessitate the gathering of more information to adequately define the perplexing situation at hand.
- **Alternative Evaluation:** The possible courses of action are enumerated and evaluated in this stage of the process. Alternatives are formulated and revaluated according to the criteria set forward by the decision maker. The alternatives may be stated either implicitly or explicitly.
- **Decision:** The final stage in the process is the selection of the alternative to address the problem identified. The decision is presumably the alternative that the decision maker finds most appealing given all the information and intuition available to him.

Decision Environment

The stages in the decision making process are by no means clearly delineated or static in nature. The various activities may be occurring simultaneously or may be so programmed that decisions are made almost instantaneously. As noted in the earlier section, the decisions and outcomes are probabilistic in nature and are characterised by some degree of uncertainty.

Role of Research in Decision Making

The role of research in decision making is relatively simple. It is to provide the decision maker with relevant and useful information with which to make decisions.

However, research is only one of four major means of obtaining information and business research is the purposeful and systematic nature of the latter. Research is specifically designed to answer managerial problems within a systematic and controlled framework so as to ensure the best possible information for decision making. This difference is noteworthy because if the research process is correctly implemented, the ultimate result should be more dependable information on which the decision maker can act.

Research and the Marketing Manager

There are a number of areas of conflict that often occur between research and management, and these conflicts end up reducing the potential effectiveness of marketing research to the organisation. While these problems are not insurmountable, they frequently require that the researcher be nearly as adept at handling human nature as he is with techniques of the research trade.

The researchers and managers have divergent backgrounds, interests and skills, as well as different organisational responsibilities. They frequently differ in terms of their appreciation for marketing research and their expectations of what research can and cannot do. The reasons are:

- (a) Intellectual differences, wherein the manager views the researcher as too academic, and too anxious to complicate rather than simplify.
- (b) A narrow conception of marketing research, where the manager views marketing research as a facts generator and not as a partner in the decision making process.
- (c) Uneven calibre of marketing researchers resulting in low-quality studies and reinforcing the management's already low view of research function, and
- (d) Late results, wherein a well-designed study may require more time than management would like.

Check Your Progress

- 8. What are strategic decisions?
- 9. Define the information search stage of decision making.

3.5 CHOOSING A GOOD RESEARCH DESIGN

Before learning about the considerations for choosing a good research design, let's learn about the components of a good research design.

A good research design consists of the following components :

- 1. *Title of the study* should be as brief and precise as possible, disclosing a brief account of the contents of the report in self-explanation.

NOTES

NOTES

2. The *problem* should state in unambiguous and more precise, clear, simple and concise terms which are most preferable.
3. A brief *survey of literature* to fill up the gaps in the construction of knowledge. It not only adds its contribution to the quantum of knowledge in the subject but also directs a correct approach in the investigation.
4. The *geographical and functional dimensions* of the study should be identified.
5. The *objectives of the study* should be laid down precisely.
6. *Hypothesis* resorting for possible reasoning is to be formulated.
7. *Methodology of investigation* is to be framed, covering data source, instruments of data collection, sampling plan and its size. Designing a research plan calls data sources, research approaches, research instruments, sampling plan and contact methods.
 - Data source* : Primary data, secondary data
 - Research approaches* : Observation, focus group survey, experiment
 - Research instruments* : Questionnaire, schedule
 - Sampling plan* : Sampling unit, sample size, sampling procedure
 - Contact methods* : Personal or telephone interview
8. Depending on the nature of the data and information required to fulfill the objective and hypothesis, data should be analyzed through appropriate *statistical tools*.
9. The *interpretation of results of an investigation, findings of the study* should objectively uncover any additional factors.
10. No research is an end by itself. The researcher must *report what he discovered during the course of investigation* in clear, specific, simple terms and directly relate to the objective of the study.

There are certain criteria that make a good research design. First, it must answer the research questions. Second, to facilitate control of independent variables, it is necessary to follow the random procedure of selection wherever possible. Third, the design should have desirable generalizability, for one would certainly like to apply the results to other situations. This is more so true in the case of applied research. This lesson describes the major types of research designs and how they can be used to reduce errors that might arise during the research process.

3.5.1 Selection of Research Design

Once the researcher has identified the research scope and objectives, he has also established his/her epistemological position. This could be positivistic—in which case the method of enquiry would necessarily be scientific and empirical. Subsequently, this would require a statistical method of analysis (Ackroyd, 1996).

The constructivists on the other hand argue for methods that are richer and more applicable to the social sciences, unlike the more pedantic experimental approach. Qualitative is a more definitive choice here than the quantitative (Atkinson and Hammersley, 1994). Yet another approach is the principle of triangulation (Jick, 1979), which advocates the simultaneous or a sequential use of the qualitative and quantitative methods of investigation. The proponents state that when the findings from diverse methods are collated, then the results are richer, more wholistic and this, in turn, improves the sanctity of the analysis.

The formulated research questions are then, through a comprehensive theoretical review, put into a practical perspective. The conceptual design thus developed requires and entails specifications of the variables under study as well as approach to the analysis. This might in turn lead to a refining or rephrasing of the defined research questions. Thus, the formulation of the research design is not a stagnant stage in the research process; rather it is an ongoing backward and forward integrated process by itself.

An illustration: Let us take the example of the organic food study. The formulated research problem was:

To investigate the consumer decision-making process for organic food products and to segment the market according to the basket size.

On conducting an extensive review of the literature, it was found that organic consumption is not always a self-driven choice; rather it could be the seller who might influence the product choice. Thus, a research design was formulated to study the organic consumer's decision stages. However, once the design is selected and a proposed sampling plan is developed, the next step required is that the constructs and the variables to be studied must be operationalized. On defining the organic consumer, we realized the significance of the psychographics of the individual—the attitude, interest and opinion—which were extremely critical. Thus, to get a wholistic view, one needs to look at the psychographic profile of the existing consumer, as well as of the potential consumer with a similar mindset. This led to a revision of the research question:

To investigate the consumer decision-making process for organic food products and to segment the market—existing and potential—according to their psychographic profile.

The researcher has a number of designs available to him for investigating the research objectives. There are various typologies that can be adopted for classifying them. You have already learnt about them in the previous section. When it comes to research design, the decision regarding qualitative and quantitative approach will help choose a research design:

Qualitative research: The design is exploratory or descriptive, loosely structured and open to interpretation and presumptions.

Quantitative research: The design is structured and has a measurable set of variables with a presumption about testing them.

NOTES

NOTES

The considerations to be made in this regard can be summarised in the following table:

- Whether the research problem is explored or not?
- Whether research literature exists or not?
- Whether variables are large and undefined or properly defined?
- Whether the researchers have comfort in unstructured or structured settings of research?
- Whether there is room for ambiguity?
- Whether researchers are willing to engage with the participants intensively or not?
- Whether description or explanation is preferred?
- Whether participant's perspectives are considered, or predictions are to be made?
- Whether the outcome is considered as a process or a product?

If the answers to the above are leaning towards the former part of the question, a qualitative research design is said to be preferred. Thus, it is to say that while choosing research design, research problem, researcher training, psychological attributes and focus of study are important bases to be considered.

Check Your Progress

- | |
|--|
| 10. What does designing a research plan call for?
11. What are the criteria that make a good research plan? |
|--|

3.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Kerlinger defines a research design as, '..... a plan, structure and strategy of investigation so conceived as to obtain answers to research questions or problems. The plan is the complete scheme or programme of the research. It includes an outline of what the investigator will do from writing the hypotheses and their operational implications to the final analysis of data.'
2. While the design is the specific framework that has been created to seek answers to the research question, the research method is the technique to collect the information required to answer the research problem, given the created framework.
3. The scenario is reproduced based upon the secondary information and a primary interview/discussion with those involved in the occurrence. Thus, there might be an element of bias as the data, in most cases, becomes a judgemental analysis rather than a simple recounting of events.

4. The approach of collecting particulars from significant and knowledgeable people is referred to as the expert opinion survey.
5. The objective of descriptive research studies is to provide a comprehensive and detailed explanation of the phenomena under study.
6. The necessary conditions for making causal inferences are: (i) concomitant variation, (ii) time order of occurrence of variables and (iii) absence of other possible causal factors.
7. The main advantages of factorial design are:
 - It is possible to measure the main effects and interaction effect of two or more independent variables at various levels.
 - It allows a saving of time and effort because all observations are employed to study the effects of each factor.
8. Strategic decisions are those that are largely unstructured and which affect the general direction of the organisation. Examples are such issues as diversification, product or market development and divestiture decisions.
9. As per the information search stage of decision making, information pertaining to the problem situation identified above is collected and assimilated into a format that is conducive to analysis of the problem at hand. Assumption and alternative scenarios are posed for subsequent problem analysis. The information generated at this stage may substantially redefine the problem initially formulated.
10. Designing a research plan calls for data sources, research approaches, research instruments, sampling plan and contact methods.
11. There are certain criteria that make a good research design. First, it must answer the research questions. Second, to facilitate control of independent variables, it is necessary to follow the random procedure of selection wherever possible. Third, the design should have desirable generalizability, for one would certainly like to apply the results to other situations.

NOTES

3.7 SUMMARY

- Once you have established the ‘what’ of the study, i.e., the research problem, the next step is the ‘how’ of the study, which specifies the method of achieving the research objectives. In other words, this is the research design.
- The researcher has a number of designs available to him for investigating the research objectives. The classification that is universally followed is the one based upon the objective or the purpose of the study.
- An exploratory design is flexible in its approach and mostly involves a qualitative investigation. The sample size is not strictly representative and at times it might only involve unstructured interviews with a couple of subject experts.

NOTES

- Secondary sources of data, as the name suggests, are data in terms of the details of previously collected findings in facts and figures—which have been authenticated and published. It is a fast and inexpensive way of collecting information.
- The case study method requires an in-depth study and is focused on a single unit of analysis. This unit could be an employee or a customer; an organization or a complete country analysis.
- In a typical focus group, there is a carefully selected small set of individuals representative of the larger respondent population under study. It is called a focus group as the selected members discuss the concerned topic for the duration of 90 minutes to, sometimes, two hours.
- Descriptive research are conclusive studies. However, they lack the precision and accuracy of experimental designs, yet it lends itself to a wide range of situations and is more frequently used in business research.
- There are also situations in which the population being studied is not of a homogeneous nature but composed of different groups. Thus it becomes essential to study the sub-segments independently. This variation of the design is termed as *multiple cross-sectional studies*.
- A distinguishing and mandatory feature of the longitudinal design is that once the sample is selected, it needs to stay constant over the period of the study. That means the number of panel members has to be the same.
- Experimental designs are conducted to infer causality. In an experiment, a researcher actively manipulates one or more causal variables and measures their effects on the dependent variables of interest.
- In true experimental designs, researchers can randomly assign test units and treatments to an experimental group. Here, the researcher is able to eliminate the effect of extraneous variables from both the experimental and control group.
- The main limitation of completely randomized designs is that it does not take into account the effect of extraneous variables on the dependent variable.
- During the exploratory research, the agency tries to identify and understand the respondents who would be interviewed in the data collection stage.
- Secondary data is generally used to understand the problem, to define the problem better, to develop an appropriate approach to the problem, to formulate a suitable research design or to answer certain questions.
- Focus groups are valuable due to the unexpected findings that emanate from free-flowing discussions among customers. Such discussions are likely to reveal complex and subtle relationships between consumers and products.

- Observation is a non-intrusive method of studying consumer behaviour and is extremely useful when the agency is unfamiliar with the research area.
- Experimental research establishes cause and effect. It involves setting up of control procedures to isolate the impact of a factor, like price discount, on a dependent variable like sales.
- *Methodology of investigation* is to be framed, covering data source, instruments of data collection, sampling plan and its size.
- The *interpretation of results of an investigation, findings of the study* should objectively uncover any additional factors.

NOTES

3.8 KEY WORDS

- **Research Design:** Research design refers to a strategy or method used to acquire the required information. It is a proper framework of procedures to be followed for collecting the information by deciding which sources are to be used.
- **Exploratory design:** It refers to the loosely structured design whose basic objective is to explore and obtain clarity about the problem situation.
- **Conclusive research design:** It is a kind of research study especially carried out to test and validate formulated hypotheses and specified relationships.
- **Descriptive research design:** It is a type of research design whose objective is to provide a comprehensive and detailed explanation of the phenomena under study.
- **Causal research:** It is a type of research design to establish the why and the how of a phenomenon. Causal research explores the effect of one thing on another and more specifically, the effect of one variable on another.

3.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. List the three basic principles that the formulated research design must ensure.
2. Mention the two essential characteristics of cross-sectional studies.
3. What are the advantages of a true panel?
4. Write a short note on external validity.
5. What does quasi-experimental design lack?
6. List the main advantages of a statistical design.

7. When is a depth interview conducted?
8. How does descriptive research differ from exploratory research?

NOTES

Long-Answer Questions

1. Analyze the essential purposes of an exploratory design study.
2. Discuss the case study method of conducting an exploratory research.
3. Explain the distinguishing features of longitudinal studies.
4. Describe the types of experimental designs.
5. Elaborate upon exploratory and secondary research.
6. Discuss the components of a good research design.

3.10 FURTHER READINGS

- Saravanel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.
- Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.
- Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.
- Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.
- Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

UNIT 4 DATA COLLECTION SOURCES AND METHODS

*Data Collection Sources
and Methods*

NOTES

Structure

- 4.0 Introduction
- 4.1 Objectives
- 4.2 Meaning and Nature of Primary and Secondary Data
- 4.3 Uses, Advantages and Disadvantages of Secondary Data
- 4.4 Types of Secondary Data Sources
- 4.5 Primary Data and its Types
- 4.6 Answers to Check Your Progress Questions
- 4.7 Summary
- 4.8 Key Words
- 4.9 Self Assessment Questions and Exercises
- 4.10 Further Readings

4.0 INTRODUCTION

A researcher needs to collect information that is study-specific. For this purpose, there exist a number of sources such as primary and secondary data. Primary data is collected directly by the researchers from main sources and they incur a high amount of time and money on these. On the other hand, secondary data is the information that has already been collected by some other researcher and is readily available to be made use of by other researchers. The advantage of this data is that it is economical to collect and easy to access and manage as it is recorded and published in a structured format. In this unit, an overview of the types, advantages and disadvantages of primary and secondary data is given.

4.1 OBJECTIVES

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

- Describe the meaning and nature of Primary and Secondary Data
- Discuss the uses, advantages and disadvantages of Secondary Data
- Analyse the types of Secondary Data Sources
- Describe Primary Data and its types

4.2 MEANING AND NATURE OF PRIMARY AND SECONDARY DATA

To understand the number of choices available to a researcher for collecting the study-specific information, one needs to be fully aware of the resources available

*Self-Instructional
Material*

NOTES

for the study and the level of accuracy required. To appreciate the truth of this statement, one needs to examine the variety of methods available to the researcher. The data sources could be either problem specific and primary or historical and secondary in nature (Figure 4.1).

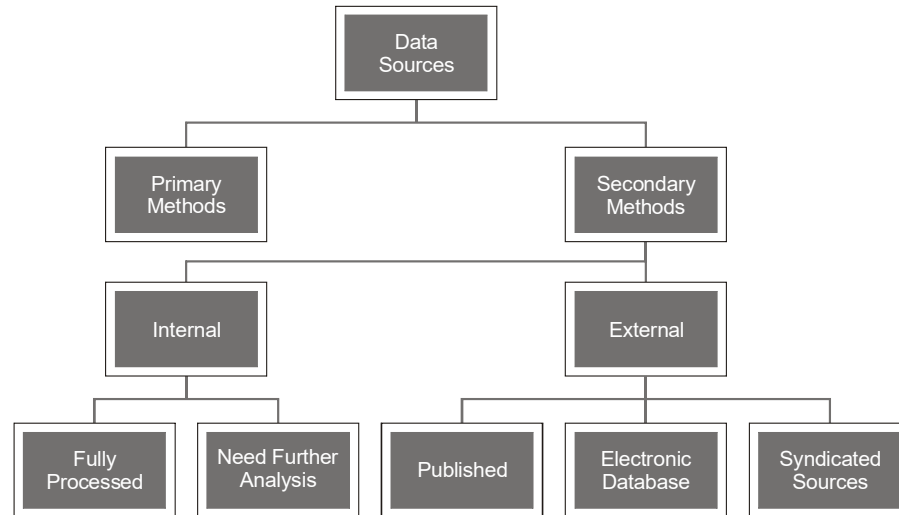


Fig. 4.1 Sources of Research Information

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

Secondary data, as the name implies, is that information which is not topical or research-specific and has been collected and compiled by some other researcher or investigative body. This type of data is recorded and published in a structured format, and thus, is quicker to access and manage. Secondly, in most instances, unless it is a data product, it is not too expensive to collect. The information required is readily available as a data product or as the audit information which the researcher or the organization can get and use it for arriving at quick decisions. In comparison to the original research-centric data, secondary data can be economically and quickly collected by the decision maker in a short span of time. However, one must remember that this is a little low on accuracy as what is primary and original for one researcher would essentially become secondary and historical for someone else.

Table 4.1 gives a snapshot of the major differences between the two methods.

Table 4.1 Primary vs Secondary Data

	Primary Data	Secondary Data
Collection purpose	For the problem at hand	For other problems
Collection process	Very involved	Rapid & easy
Collection cost	High	Relatively low
Collection time	Long	Short
Accuracy	As Desired	Not Known
Control	Complete	No Control
Period	Up to date	Dated

NOTES

Check Your Progress

1. Which out of the following data are considered to be problem specific: primary or secondary data?
2. Why is secondary data quicker to manage?

4.3 USES, ADVANTAGES AND DISADVANTAGES OF SECONDARY DATA

We have already discussed what secondary data is. Let's now study its uses, advantages and disadvantages.

Uses of Secondary Data

Secondary data can be used for multiple purposes as follows:

- **Problem identification and formulation stage:** Existing information on the topic under study is useful to help develop the research question.
- **Hypotheses designing:** Previous research studies done in the area could help in hypothesizing about expected results.
- **Sampling considerations:** There might be respondent related databases available to seek respondent statistics and relevant contact details. These would help during sampling for the study.
- **Primary base:** The secondary information collected can be used to design the primary data collection instruments, in order to phrase and design the right questions.
- **Validation board:** Earlier records and studies can also be used to support or validate the information collected through primary sources.

Before we examine the wide range of the secondary sources available to the business researcher, it is essential that one is aware of the advantages and disadvantages of using secondary sources.

NOTES

Advantages and Disadvantages of Secondary Data

There are multiple advantages of using secondary data.

- **Resource advantage:** Any research that is making use of secondary information will be able to save immensely in terms of both cost and time
- **Accessibility of data:** The other major advantage of secondary sources is that it is very easy to access this data.
- **Accuracy and stability of data:** Data from recognized sources has the additional advantage of accuracy and reliability
- **Assessment of data:** It can be used to compare and support the primary research findings of the present study.

However, there is need for caution as well because in using secondary data, there might be some disadvantages like:

- **Applicability of data:** The information might not be directly suitable for our study. Also since it is past data it might not be applicable today.
- **Accuracy of data:** All data that is available might not be reliable and accurate.

Check Your Progress

3. What can secondary information be used for?
4. List the disadvantages of using secondary data.

4.4 TYPES OF SECONDARY DATA SOURCES

As we saw earlier in Figure 4.1, secondary data can be divided into internal and external sources. Internal, as the name implies, is organization-or environment-specific source and includes the historical output and records available with the organization which might be the backdrop of the study. The data that is independent of the organization and covers the larger industry-scape would be available in the form of published material, computerized databases or data compiled by syndicated services. Discussed below are three major sources of data – internal, external, computer-stored data and syndicated databases.

1. Internal sources of data

Compilation of various kinds of information and data is mandatory for any organization that exists. Some sources of internal information are presented in Figure 4.2.

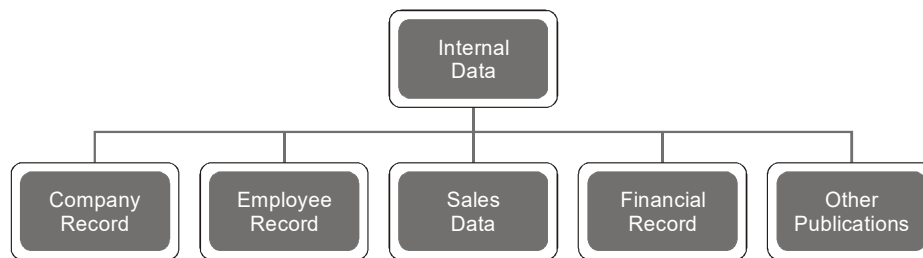


Fig. 4.2 Internal Sources of Data

NOTES

- **Company records:** This includes all the data about the inception, the owners, and the mission and vision statements, infrastructure and other details, including both the process and manufacturing (if any) and sales, as well as a historical timeline of the events.
- **Employee records:** All details regarding the employees (regular and part-time) would be part of employee records.
- **Sales data:** This data can take on different forms:
 - (i) **Cash register receipt**
 - (ii) **Salespersons' call records:** This is a document to be prepared and updated every day by each individual salesperson.
 - (iii) **Sales invoices:** Customer who has placed an order with the company, his complete details including the size of the order, location, price by unit, terms of sale and shipment details (if any).
- **Financial records and sales reports**

Besides this, there are other published sources like warranty records, CRM data and customer grievance data which are extremely critical in evaluating the health of a product or an organization.

2. External data sources

As stated earlier, information that is collected and compiled by an outside source that is external to the organization is referred to as *external* source of data. External sources of data include the following:

Published data: There could be two kinds of published data—one that is from the official and government sources and the other kind of data is that which has been prepared by individuals or private agencies or organizations.

Government sources: The Indian government publishes a lot of documents that are readily available and are extremely useful for the purpose of providing background data. A brief snapshot of some government data is given in Table 4.2.

Table 4.2 Secondary Data—Government Publications

NOTES

	Sub-type	Sources	Data	Uses
1.	Census data conducted every ten years throughout the country	Registrar General of India conducting census survey http://censusindia.gov.in/	Size of the population and its distribution by age, sex, occupation and income levels. 2010 census is taking many more variables to get a better picture of the population	Population information is significant as the forecasts of purchase, estimates of growth and development, as well as policy decisions can be made on this base
2.	Statistical Abstract India – annually	CSO (Central Statistical Organization) for the past 5 years http://www.mospi.gov.in/cso_test1.htm	Education, health, residential information at the state level is part of this document	Making demand, estimations and a state level assessment of government support and policy changes can be made
3.	White paper on national income	CSO http://www.mospi.gov.in/cso_test1.htm	Estimates of national income, savings and consumption	Significant indication of the financial trends; investment forecasts and monetary policy formulation
4.	Annual Survey of Industries – all industries	CSO No. of units, persons employed, capital output ratio, turnover, etc. http://www.mospi.gov.in/cso_test1.htm		Information on existing units gives perspective on the Industrial development and helps in creating the employee profile
5.	Monthly survey of selected industries	CSO http://www.mospi.gov.in/cso_test1.htm	Production statistics in detail	Demand–supply estimations
6.	Foreign Trade of	Director General of Commercial Intelligence	Exports and Imports	Forecast, manufacturing

7.	Wholesale price index—weekly all India Consumer Price Index	Ministry of Commerce and Industry http://india.gov.in/sectors/commerce/ministry_commerce.php	Reporting of prices of products like food articles, foodgrains, minerals, fuel, power, lights, lubricants, textiles, chemicals, metal, machinery and transport	Establishing price bands of product categories; pricing estimations for new products; determining consumer spend
8.	Economic Survey – annual publication	Dept. of Economic Affairs, Ministry of Finance, patterns, currency and finance http://finmin.nic.in/the_ministry/dept_eco_affairs/	Descriptive reporting of the current economic status	Estimations of the future and evaluation of policy decisions and extraneous factors in that period
9.	National Sample Survey (NSS)	Ministry of Planning http://www.planningcommission.gov.in/	Social, economic, demographic, industrial and agricultural statistics.	Significant for making policy decisions as well as studying sociological patterns

Data Collection Sources and Methods

NOTES

Other data sources: This source is the most voluminous and most frequently used, in every research study. The information could be

- *Books and periodicals*
- *Guides:* including Industry guides
- *Directories and indices*
- *Standard non-governmental statistical data:* Some non-government data sources are presented in Table 4.3.

Table 4.3 Secondary Data—Non-government Publications

	Sub-type	Sources	Data	Uses
1.	Company Working Results – Stock Exchange Directory	Bombay Stock Exchange http://www.bseindia.com/	A complete database of the companies registered with the stock exchange and comprehensive details about stock policies and current share prices	Significant in determining the financial health of various sectors as well as assessment of corporate funding and predictions of outcomes

NOTES

2.	Status reports by various commodity boards	The commodity board or the industry associations like Jute Board, Cotton Industry, Sugar Association, Pulses Board, Metal Board, Chemicals, Spices, Fertilizers, Coir, Pesticides, Rubber, Handicrafts, Plantation Boards, etc.	Detailed information on current assets-in terms of units, current production figures and market condition	These are useful for individual sectors in working out their plans as well as evaluating causes of success or failure
3.	Industry Associations on problems faced by private sector, etc.	FICCI, ASSOCHAM, AIMA, Association of Chartered Accountants and Financial Analysts, Indo-American Chamber of Commerce, etc. http://www.ficci.com/ http://www.assocham.org/ http://www.aima-ind.org/ www.iaccindia.com/	Cases/ comprehensive reports by the supplier or user or any other section associated with the sector	Cognizance of the gaps and problems in the effective functioning of the organization; trouble shooting
4.	Export related data – commodity wise	Leather Exports Promotion Council, Apparel Export Promotion Council, Handicrafts, Spices Tea, etc., Exim Bank etc. http://www.leatherindia.org/ http://www.aepcindia.com/	Product and country wise data on the export figures as well as information on existing policies related to the sector	To estimate the demand; gauge opportunities for trade and impetus required in terms of manufacturing and policy changes
5.	Retail Store Audit on pharmaceutical, veterinary, consumer products	ORG (Operations Research Group); Monthly reports on urban sector. Quarterly reports on rural sector	The touch point for this data is retailer, who provides the figures related to product sales; the data is very comprehensive and covers most brands. The data is region specific and covers both inventory and goods sold	Market analysis and market structure mapping with estimations of market share of leading brands. The audit can also be used to study consumption trends at different time periods or subsequent to sales promotion or other activities

NOTES

6.	National Reader-ship Surveys (NRS)	IMRB-survey of reading behaviour for different segments as well as different products http://www.imrbint.com/	Today these surveys are done by various bodies with different sample bases. Today the survey base has become younger, with the age of the reader lowered to 12+	Media planning and measuring exposure as well as reach for product categories
7.	THOMPS ON INDICES: Urban market index, Rural market index	Hindustan Thompson Associates	All towns with population of more than one lakh are covered and information of demographic and socio-economic variables are given for each city with Mumbai as base. The rural index similarly covers about 400 districts with socio-economic indicators like value of agriculture output, etc.	The inclinations to purchase consumer products are directly related to socio-economic development of communities in general. The indices provide barometers to measure such potentials for each city and has implications for the researcher in terms of data collection sources

3. Computer-stored data

Information today is also available in an electronic form. The databases available to the researcher can be classified on the basis of the type of information or by the method of storage and recovery as described below. Figure 4.3 gives a classification of the sources of computerized data.

- **Reference databases:** These refer users to the articles, research papers, abstracts and other printed news contained in other sources. They provide online indices and abstracts and are thus also called bibliographic databases.
- **Source databases:** These provide numerical data, complete text, or a combination of both.

Based on storage and recovery mechanisms: Another useful way of classifying databases is based on their method of storage and retrieval.

- **Online databases:** These can be accessed in real time directly from the producers of the database or through a vendor. Examples include ABI/Inform, EBSCO and Emerald.

NOTES

- **CD-ROM databases:** Here information is available on a CD-ROM.

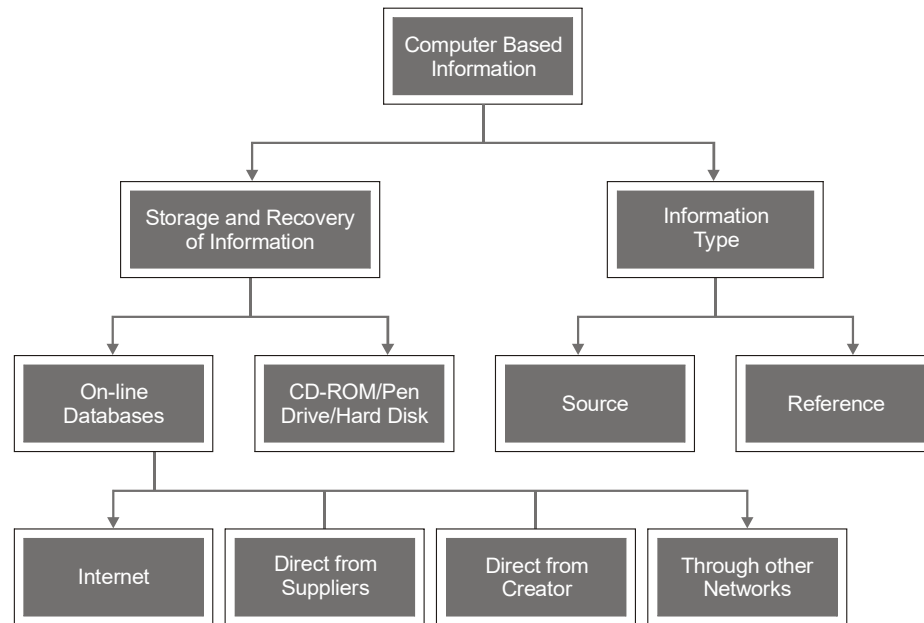


Fig. 4.3 Classifications of Computerized Databases

4. Syndicated data sources

Syndicated service agencies are organizations that collect organization/product-category-specific data from a regular consumer base and create a common pool of data that can be used by multiple buyers, for their individual purpose.

There are different ways to classify syndicate sources.

- **Household/individual data:** These could be in the form of surveys or panel data available through reputed agencies.
- **Surveys:** Surveys are usually one-time assessments conducted on a large representative respondent base. For example, opinion polls before elections, best business school to study etc.
- **Product purchase panels:** These specially selected respondent groups specifically record certain identified purchases, generally related to household products and groceries.
- **Media-specific panels:** Panels are also created for collecting information related to promotion and advertising. The task of the media panel is to make use of different kinds of electronic equipment to automatically record consumer viewing behaviour. These are used to calculate the television rating performance (TRP) of different programs.
- **Scanner devices and individual source systems:** To overcome the problems of panel data, a new service is provided by research agencies through electronic scanner devices-e.g. sales volume tracking data.

- **Institutional syndicated data:** The syndicated data can also be available at the institutional level. Retailer and wholesaler audits are examples of this kind. Usually the records are noted as:

Beginning stocks + deliveries – ending inventory = sales for the period

NOTES

Check Your Progress

5. What is internal source of secondary data?
6. What do company records include?
7. Why is population information significant?
8. What information does the economic survey provide?
9. Name some online databases.

4.5 PRIMARY DATA AND ITS TYPES

The researcher has available to him/her a wide variety of data collection methods which are primary or problem specific in nature. However in this section, we will be discussing the major and most often used methods like the observation method, focus group discussion and interview method. The questionnaire method is the most commonly used method of primary data collection. This will be dealt with separately in Unit 5. Let us discuss some of the other widely used methods now.

(i) Observation

Observation is a direct method of collecting primary data. It is one of the most appropriate methods to use in case of descriptive research. The method of observation involves viewing and recording individuals, groups, organizations or events in a scientific manner in order to collect valuable data related to the topic under study.

The mode of observation could be in a standardized or *structured observation*. Here, the nature of content to be recorded, the format and the broad areas of recording are predetermined. Thus, the observer's bias is reduced and the authenticity and reliability of the information collected is higher. For example, Fisher Price toys carry out an observational study whenever they come out with a new toy. The observer is supposed to record the appeal of the toy for a child.

The opposite of this is called the *unstructured observation*. Here, the observer is supposed to make a note of whatever he understands as relevant for the research study. This kind of approach is more useful in exploratory studies. Since it lacks structure, the chances of observer's bias are high. An example of this is the observation of consumers at a bank, a restaurant or a doctor's clinic.

However, it is critical here to understand that the researcher must have a preconceived plan to capture the observations made. It is not to be treated as a

NOTES

blank sheet where the observer reports what he sees. The aspects to be observed must be clearly listed as in an audit form, or they could be indicative areas on which the observation is to be made.

Another way of distinguishing observations is the level of respondent being aware of being observed or not. This might be disguised; here the observation is done without the respondent's knowledge who has no idea that he/she is being observed. This can also be done with devices like a one-way mirror or a hidden camera or a recorder. The only disadvantage is this is ethically an intrusion of an individual's right to privacy. On the other hand, the knowledge that the person is under observation can be conveyed to the respondent, and this is undisguised observation. The decision to choose one over the other depends upon the nature of the study.

The observation method can also be distinguished on the basis of the setting in which the information is being collected. This could be natural observation, which, as the name suggests, is carried out in actual real life locations, for example the observations of how employees interact with each other during lunch breaks. On the other hand, it could be an artificial or simulated environment in which the respondent is to be observed. This is actively done in the armed forces where stress tests are carried out to measure an individual's tolerance level.

There is another differentiation where the observation could be done by a human observer or a mechanical device.

Human observation: As the name suggests, this technique involves observation and recording done by human observers. The task of the observer is simple and predefined in case of a structured observation study as the format and the areas to be observed and recorded are clearly defined. In an unstructured observation, the observer records in a narrative form the entire event that he has observed.

Mechanical observation: In these methods, man is replaced by machine. Some examples are

- Store cameras and cameras in banks and other service areas.
- Universal product code (UPC) scanned by electric scanners in stores.
- Psychogalvanometer, which measures galvanic skin response (GSR) or changes in the electrical resistance of the skin. Thus, the respondent could be exposed to different kinds of packaging, advertisements and product composition, to note his/her reaction to them.
- Eye-tracking equipment such as oculometers, eye cameras or eye view minuters, record the movements of the eye. The oculometer determines what the individual is looking at, while the pupilometer measures the interest of the person in the stimulus. The pupilometer measures changes in the diameter of the respondent's pupils.
- Trace analysis; in this the remains or the leftovers of the consumers' basket—like his credit card spend, his recycle bin on his computer, his garbage

(garbology) are evaluated to measure current trends and patterns of usage and disposal.

Observational techniques are an extremely useful method of primary data collection and are always a part of the inputs, whether accompanying other techniques, like interviews, discussions or questionnaire administration, or as the prime method of data collection. However, the disadvantage which they suffer from is that they are always behaviourally driven and cannot be used to investigate the reasons or causes of the observed behaviour. Another problem is that if one is observing the occurrence of a certain phenomena, one has to wait for the event to occur. One alternative to this is to study the recordings, whether verbal, written or audio-visual, in order to formulate the study-related inferences.

(ii) Focus Group Discussion

Focus group discussion (FGD) is a highly versatile and dynamic method of collecting primary data from a representative group of respondents. The process generally involves a moderator who steers the discussion on the topic under study. There are a group of carefully selected respondents who are specifically invited and gathered at a neutral setting. The moderator initiates the discussion and then the group carries it forward by holding a focused and an interactive discussion.

Key Elements of a Focus Group

- **Size:** Ideal recommended size for a group discussion is 8 to 12 members. Less than eight would not generate all the possible perspectives on the topic and the group dynamics required for a meaningful session. And more than 12 would make it difficult to get any meaningful insight.
- **Nature:** Individuals who are from a similar background—in terms of demographic and psychographic traits—must be included, otherwise disagreement might emerge as a result of other factors rather than the one under study. The other requirement is that the respondents must be similar in terms of the subject/policy/product knowledge and experience with the product under study. Moreover, the conduction of the focus group discussion must ensure that the following criteria are taken care of:
- **Acquaintance:** It has been found that knowing each other in a group discussion is disruptive and hampers the free flow of the discussion. It is recommended that the group should consist of strangers rather than subjects who know each other.
- **Setting:** The space or setting in which the discussion takes place should be as neutral, informal and comfortable as possible. In case one-way mirrors or cameras are installed, there is a need to ensure that these gadgets are not directly visible.
- **Time period:** The discussion should be held in a single setting unless there is a 'before' and 'after' design, which requires group perceptions, initially

NOTES

NOTES

before the study variable is introduced; and later in order to gauge the group's reactions. The ideal duration of conduction should not exceed an hour and a half. This is usually preceded by a short rapport formation session between the moderator and the group members.

- **The recording:** This is most often machine recording even though sometimes this may be accompanied by human recording as well.
- **The moderator:** The moderator is the one who manages the discussion. He might be a participant in the group discussion or he might be a non-participant. He must be a good listener and unbiased in his conduct of the discussions.

Steps in planning and conducting focus groups

The focus group conduction has to be done in a stepwise manner:

- Clearly define and enlist the research objectives of the study that requires group discussion.
- A comprehensive moderator's structured outline for conducting the whole process needs to be charted out.
- After this, the actual focus group discussion is carried out.
- The focus summary of the findings are clubbed under different heads as indicated in the focus group objectives and reported in a narrative form. This may include expressions like 'majority of the participants were of the view' or 'there was a considerable disagreement on this issue'.

Types of focus groups

The researcher has different kinds of group discussion methods available to him or her. These are:

- **Two-way focus group:** Here one respondent group sits and listens to the other and after learning from them or understanding the needs of the group, they carry out a discussion amongst themselves. For example, in a management school, the faculty group could listen to the opinions and needs of the student group.
- **Dual-moderator group:** Here, there are two different moderators; one responsible for the overt task of managing the group discussion and the other for the second objective of managing the 'group mind' in order to maximize the group performance.
- **Fencing-moderator group:** The two moderators take opposite sides on the topic being discussed and thus, in the short time available, ensure that all possible perspectives are thoroughly explored.
- **Friendship groups:** There are situations where the comfort level of the members needs to be high so that they elicit meaningful responses. This is especially the case when a supportive peer group encourages admission about the related organizations or people/issues.

- **Mini-groups:** These groups might be of a smaller size (usually four to six) and are usually expert groups/committees that on account of their composition are able to decisively contribute to the topic under study.
- **Creativity group:** These are usually longer than one and a half hour duration and might take the workshop mode. Here, the entire group is instructed, after which they brainstorm in smaller sub-groups. They then reassemble to present their sub-group's opinion. This might also stretch across a day or two.
- **Brand-obsessive group:** These are special respondent sub-strata who are passionately involved with a brand or product category (say, cars). They are selected, as they can provide valuable insights that can be successfully incorporated into the brand's marketing strategy.
- **Online focus group:** This is a recent addition to the methodology and is extensively used today. Here, the respondents at the designated time in a web-based chat room and enters their ID and password to log on. The discussion between the moderator and the participants is real time.

NOTES

(iii) Personal Interview Method

Personal interview is a one-to-one interaction between the investigator/interviewer and the interviewee. The purpose of the dialogue is research specific and ranges from completely unstructured to highly structured.

Uses of the Interview Method

The interview has varied applications in business research and can be used effectively in various stages.

- **Problem definition:** The interview method can be used right in the beginning of the study. Here, the researcher uses the method to get a better clarity about the topic under study.
- **Exploratory research:** This method can be actively used here because the structure is loose.
- **Primary data collection:** There are situations when the method is used as a primary method of data collection, this is generally the case when the area to be investigated is high on emotional responses

The Interview Process

The steps undertaken for the conduction of a personal interview are somewhat similar in nature to those of a focus group discussion.

- Interview objective:** The information needs that are to be addressed by the instrument should be clearly spelt out as study objectives. This step includes a clear definition of the construct/variable(s) to be studied.
- Interview guidelines:** A typical interview may take from 20 minutes to close to an hour. A brief outline to be used by the investigator is formulated depending upon the contours of the interview.

NOTES

(c) **Structure:** Based on the needs of the study, the actual interview may be unstructured, semi-structured or structured.

- *Unstructured:* This type of interview has no defined guidelines. It usually begins with a casually worded opening remark like ‘so tell us/me something about yourself’. The direction the interview will take is not known to the researcher also. The probability of subjectivity is very high.
- *Semi-structured:* This has a more defined format and usually only the broad areas to be investigated are formulated. The questions, sequence and language are left to the investigator’s choice. Probing is of critical importance in obtaining meaningful responses and uncovering hidden issues. After asking the initial question, the direction of the interview is determined by the respondent’s initial reply, the interviewer’s probes for elaboration and the respondent’s answers.
- *Structured:* This format has highest reliability and validity. There is considerable structure to the questions and the questioning is also done on the basis of a prescribed sequence. They are sometimes used as the primary data collection instrument also.

(d) **Interviewing skills:** The quality of the output and the depth of information collected depend upon the probing and listening skills of the interviewer. His attitude needs to be as objective as possible.

(e) **Analysis and interpretation:** The information collected is not subjected to any statistical analysis. Mostly the data is in narrative form, in the case of structured interviews it might be summarized in prose form.

Types of Interviews

There are various kinds of interview methods available to the researcher. In the last section we have spoken about a distinction based on the level of structure. The other classification is based on the mode of administering the interview.

Figure 4.4 presents a classification of the types of personal interview.

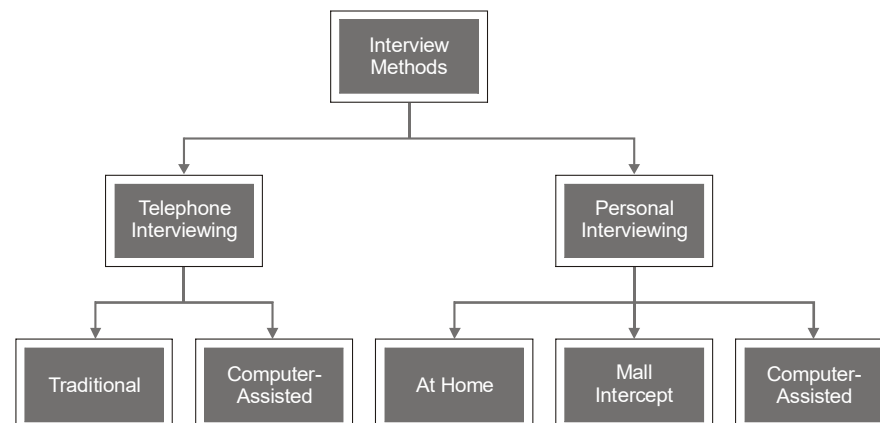


Fig. 4.4 Types of Personal Interview

(a) **Personal methods:** These are the traditional one-to-one methods that have been used actively in all branches of social sciences. However, they are distinguished in terms of the place of conduction.

- *At-home interviews:* This face-to-face interaction takes place at the respondent's residence. Thus, the interviewer needs to initially contact the respondent to ascertain the interview time.
- *Mall-intercept interviews:* As the name suggests, this method involves conducting interviews with the respondents as they are shopping in malls. Sometimes, product testing or product reactions can be carried out through structured methods and followed by 20–30 minute interviews to test the reactions.
- *Computer-assisted personal interviewing (CAPI):* This technique is carried out with the help of the computer. In this form of interviewing, the respondent faces an assigned computer terminal and answers a questionnaire on the computer screen by using the keyboard or a mouse. A number of pre-designed packages are available to help the researcher design simple questions that are self-explanatory and instead of probing, the respondent is guided to a set of questions depending on the answer given. There is usually an interviewer present at the time of respondent's computer-assisted interview and is available for help and guidance, if required.

(b) **Telephone method:** The telephone method replaces the face-to-face interaction between the interviewer and interviewee, by calling up the subjects and asking them a set of questions. The advantage of the method is that geographic boundaries are not a constraint and the interview can be conducted at the individual respondent's location. The format and sequencing of the questions remains the same.

- *Traditional telephone interviews:* The process can be accomplished using the traditional telephone for conducting the questioning.
- *Computer-assisted telephone interviewing:* In this process, the interviewer is replaced by the computer and it involves conducting the telephonic interview using a computerized interview format. The interviewer sits in front of a computer terminal and wears a mini-headset, in order to hear the respondent answer. However, unlike the traditional method where he had to manually record the responses, the responses are simultaneously recorded on the computer.

Since the interview requires a one-to-one dialogue to be carried out, it is more cumbersome and costly as compared to a focus group discussion. Also, conduction of interview requires considerable skills on part of the interviewer and thus adequate training in interviewing skills is needed for capturing a comprehensive study-related data.

NOTES

NOTES

Check Your Progress

10. Why is the authenticity of data collected by a structured observation higher?
11. What is the disadvantage of observational techniques?
12. What happens in an online focus group?

4.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Primary data, as the name suggests, is original, problem- or project-specific and collected for the specific objectives and needs spelt out by the researcher. The accuracy and relevance is reasonably high. The time and money required for this are also quite high
2. Secondary data is recorded and published in a structured format, and thus, is quicker to access and manage.
3. The secondary information collected can be used to design the primary data collection instruments, in order to phrase and design the right questions.
4. The disadvantages of using secondary data are:
 - Applicability of data: The information might not be directly suitable for our study. Also since it is past data, it might not be applicable today.
 - Accuracy of data: All data that is available might not be reliable and accurate.
5. Internal source of secondary data is organization-or environment-specific source and includes the historical output and records available with the organization which might be the backdrop of the study.
6. Company records include all the data about the inception, the owners, and the mission and vision statements, infrastructure and other details, including both the process and manufacturing (if any) and sales, as well as a historical timeline of the events.
7. Population information is significant as the forecasts of purchase, estimates of growth and development, as well as policy decisions can be made on this base.
8. The economic survey provides estimations of the future and evaluation of policy decisions and extraneous factors in that period.
9. Some online databases are ABI/Inform, EBSCO and Emerald.
10. In structured observation, the nature of content to be recorded, the format and the broad areas of recording are predetermined. Thus, the observer's bias is reduced and the authenticity and reliability of the information collected is higher.

11. The disadvantage that observational techniques suffer from is that they are always behaviourally driven and cannot be used to investigate the reasons or causes of the observed behaviour.
12. In an online focus group, the respondents at the designated time in a web-based chat room and enter their ID and password to log on. The discussion between the moderator and the participants is real time.

NOTES

4.7 SUMMARY

- Primary data, as the name suggests, is original, problem- or project-specific and collected for the specific objectives and needs spelt out by the researcher.
- Any research that is making use of secondary information will be able to save immensely in terms of both cost and time.
- Secondary data can be divided into internal and external sources. Internal, as the name implies, is organization-or environment-specific source and includes the historical output and records available with the organization which might be the backdrop of the study.
- There could be two kinds of published data—one that is from the official and government sources and the other kind of data is that which has been prepared by individuals or private agencies or organizations.
- Reference databases refer users to the articles, research papers, abstracts and other printed news contained in other sources. They provide online indices and abstracts and are thus also called bibliographic databases.
- Syndicated service agencies are organizations that collect organization/ product-category-specific data from a regular consumer base and create a common pool of data that can be used by multiple buyers, for their individual purpose.
- To overcome the problems of panel data, a new service is provided by research agencies through electronic scanner devices-e.g. sales volume tracking data.
- The method of observation involves viewing and recording individuals, groups, organizations or events in a scientific manner in order to collect valuable data related to the topic under study.
- The observation method can also be distinguished on the basis of the setting in which the information is being collected. This could be natural observation, which is carried out in actual real life locations. On the other hand, it could be an artificial or simulated environment in which the respondent is to be observed.
- The task of the observer is simple and predefined in case of a structured observation study as the format and the areas to be observed and recorded

NOTES

- are clearly defined. In an unstructured observation, the observer records in a narrative form the entire event that he has observed.
- Focus group discussion (FGD) is a highly versatile and dynamic method of collecting primary data from a representative group of respondents. The process generally involves a moderator who steers the discussion on the topic under study.
 - The focus summary of the findings are clubbed under different heads as indicated in the focus group objectives and reported in a narrative form.
 - In a fencing moderator group, two moderators take opposite sides on the topic being discussed and thus, in the short time available, ensure that all possible perspectives are thoroughly explored.
 - The interview method can be used right in the beginning of the study. Here, the researcher uses the method to get a better clarity about the topic under study.
 - At-home interview is a face-to-face interaction takes place at the respondent's residence. Thus, the interviewer needs to initially contact the respondent to ascertain the interview time.
 - Computer-assisted personal interviewing (CAPI) This technique is carried out with the help of the computer. In this form of interviewing, the respondent faces an assigned computer terminal and answers a questionnaire on the computer screen by using the keyboard or a mouse.
 - The telephone method replaces the face-to-face interaction between the interviewer and interviewee, by calling up the subjects and asking them a set of questions. The advantage of the method is that geographic boundaries are not a constraint and the interview can be conducted at the individual respondent's location.

4.8 KEY WORDS

- **Secondary Data:** It refers to the information which is not topical or research-specific and has been collected and compiled by some other researcher or investigative body. This type of data is recorded and published in a structured format, and thus, is quicker to access and manage.
- **External source of Data:** It is information that is collected and compiled by an outside source that is external to the organization.
- **Syndicated Service Agencies:** These are organizations that collect organization/product-category-specific data from a regular consumer base and create a common pool of data that can be used by multiple buyers, for their individual purpose.

- **Galvanic Skin Response (GSR):** It refers to changes in sweat gland activity that are reflective of the intensity of our emotional state, otherwise known as emotional arousal.
- **Computer-Assisted Personal Interview (CAPI):** It is a face-to-face data collection method in which the interviewer uses a tablet, mobile phone or a computer to record answers given during the interview.

NOTES

4.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. State the advantages of using secondary data.
2. Write a short note on the forms of sales data.
3. What are the two types of published data?
4. What are status reports by various commodity boards useful for?
5. What is the use of export related data?
6. What type of information do source databases provide?
7. What kind of individuals should be included in a focus group?
8. Write a short note on two-way focus group.

Long-Answer Questions

1. Discuss the purposes for which secondary data can be used.
2. Analyse some government sources of data.
3. Explain the different ways to classify syndicate sources.
4. Analyse some examples of mechanical observation.
5. Describe the key elements of a focus group.

4.10 FURTHER READINGS

- Saravanel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.
- Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.
- Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.
- Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.
- Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

BLOCK - II

RESEARCH METHODOLOGY

NOTES

UNIT 5 DESIGNING QUESTIONNAIRE

Structure

- 5.0 Introduction
- 5.1 Objectives
- 5.2 Questionnaire Building
- 5.3 Process of Questionnaire Design
- 5.4 Answers to Check Your Progress Questions
- 5.5 Summary
- 5.6 Key Words
- 5.7 Self Assessment Questions and Exercises
- 5.8 Further Readings

5.0 INTRODUCTION

A Questionnaire is a research technique used for obtaining useful statistical information by asking questions to respondents. Before asking questions, a series of questions is prepared as per the objective and purpose of the survey. This technique is a method of collecting primary data in a feasible way. Subsequently, a questionnaire could be formalized, concealed, non-formalized and unconcealed. The process of developing a questionnaire design follows a systematic series of steps that includes determining the content of the questionnaire and motivating the respondents to answer correctly among a number of other significant steps. In this unit, the process of questionnaire building and questionnaire design have been focused upon.

5.1 OBJECTIVES

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

- Describe the concept and criteria of Questionnaire Building
- Explain the types of questionnaire
- Discuss the process of Questionnaire Design

5.2 QUESTIONNAIRE BUILDING

The questionnaire is a research technique that consists of a series of questions asked to respondents, in order to obtain statistically useful information about a

given topic. It is one of the most cost-effective methods of collecting primary data, which can be used with considerable ease by most individual and business researchers. It has the advantage of flexibility of approach and can be successfully adapted for most research studies. The instrument has been defined differently by various researchers. Some take the traditional view of a written document requiring the subject to record his/her own responses (Kervin, 1999). Others have taken a broader perspective to include structured interview also as a questionnaire (Bell, 1999). It is essentially a data-collection instrument that has a predesigned set of questions, following a particular structure (De Vaus, 2002). Since it includes a standard set of questions, it can be successfully used to collect information from a large sample in a reasonably short time period.

However, the use of questionnaire is not always the best method in all research studies. For example, at the exploratory stage, rather than questionnaire, it is advisable to use a more unstructured interview. Secondly, when the number of respondents is small and one has to collect more subjective data, then a questionnaire is not advisable.

Criteria for Designing a Questionnaire

There are certain criteria that must be kept in mind while designing the questionnaire. The first and foremost requirement is that the spelt-out research objectives must be converted into clear questions which will extract answers from the respondent. This is not as easy as it sounds, for example, if one wants to know how many times your teacher praised you in the week? It is very difficult to give an exact number. The second requirement is, it should be designed to engage the respondent and encourage a meaningful response. For example, a questionnaire measuring stress cannot have a voluminous set of questions which fatigue the subject. The questions, thus, should encourage response and be easy to understand. Lastly, the questions should be self-explanatory and not confusing as then the person will answer the way he understood the question and not in terms of what was asked. This will be discussed in detail later, when we discuss the wording of the questions.

Types of Questionnaire

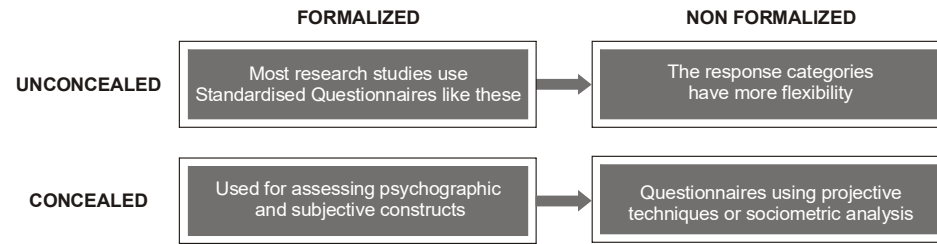
There are different types of questionnaires available to the researcher. The categorization can be done on the basis of a variety of parameters. The two criteria that are most frequently used for designing purposes are the degree of structure and the degree of concealment. Structure refers to the degree to which the response category has been defined. Concealment refers to the degree to which the purpose of the study is explained to the respondent.

Instead of considering them as individual types, most research studies use a mixed format. Thus, they will be discussed here as a two-by-two matrix (Table 5.1).

NOTES

Table 5.1 Types of Questionnaire

NOTES



Let us discuss the types of questionnaires. Questionnaires can be categorized on the basis of their structure or method of administration.

Based on the structure, questionnaires can be divided into the following categories:

Formalized and unconcealed questionnaire: This is the one that is the most frequently used by all management researchers. For example, if a new brokerage firm wants to understand the investment behaviour of people, they would structure the questions and answers as follows:

1. Do you carry out any investment(s)?
Yes _____ No _____
If yes, continue, else terminate.
2. Out of the following options, where do you invest? (tick all that apply).
Precious metals _____, real estate _____, stocks _____,
government instruments _____, mutual funds _____,
any other _____.

This kind of structured questionnaire is easy to administer, and has both the questions as self-explanatory and the answer categories clearly defined.

Formalized and concealed questionnaire: These questionnaires have a formal method of questioning; however the purpose is not clear to the respondent. The research studies which are trying to find out the latent causes of behaviour and cannot rely on direct questions use these. For example young people cannot be asked direct questions on whether they are likely to indulge in corruption at work. Thus, the respondent has to be given a set of questions that can give an indication of what are his basic values, opinions and beliefs, as these would influence how he would react to issues.

Non-formalized and unconcealed: Some researchers argue that rather than giving the respondents pre-designed response categories, it is better to give them unstructured questions where they have the freedom of expressing themselves the way they want to. Some examples of these kinds of questions are given below:

1. Why do you think Maggi noodles are liked by young children?

2. How do you generally decide on where you are going to invest your money?

3. Give THREE reasons why you believe that the show *Satyamev Jayate* has affected the common Indian person?

The data obtained here is rich in content, but quantification cannot go beyond frequency and percentages to represent the findings.

Non-formalized and concealed: If the objective of the research study is to uncover socially unacceptable desires and subconscious and unconscious motivations, the investigator makes use of questions of low structure and disguised purpose. However, these require interpretation that is highly skilled. Cost, time and effort are also much higher than others.

Another useful way of categorizing questionnaires is on the method of administration. Thus, the questionnaire that has been prepared would necessitate a face-to-face interaction. In this case, the interviewer reads out each question and makes a note of the respondent's answers. This administration is called a *schedule*. It might have a mix of the questionnaire type as described in the section above and might have some structured and some unstructured questions. The other kind is the *self-administered questionnaire*, where the respondent reads all the instructions and questions on his own and records his own statements or responses. Thus, all the questions and instructions need to be explicit and self-explanatory.

Another useful way of categorizing questionnaires is on the method of administration. Thus, the questionnaire that has been prepared would necessitate a face-to-face interaction. In this case, the interviewer reads out each question and makes a note of the respondent's answers. This administration is called a *schedule*. It might have a mix of the questionnaire type as described in the section above and might have some structured and some unstructured questions. The investigator might also have a set of additional material like product prototypes or copy of advertisements. The investigator might also have a predetermined set of standardized questions or clarifications, which he can use to ask questions like 'why do you say that?' or 'can you explain this in detail' 'what I mean to ask is.....' The other kind is the *self-administered questionnaire*, where the respondent reads all the instructions and questions on his own and records his own statements or responses. Thus, all the questions and instructions need to be explicit and self-explanatory.

The selection of one over the other depends on certain study prerequisites.

Population characteristics: In case the population is illiterate or unable to write the responses, then one must as a rule use the schedule, as the questionnaire cannot be effectively answered by the subject himself.

Population spread: In case the sample to be studied is large and widely spread, then one needs to use the questionnaire. When the resources available for the study are limited, then schedules become expensive to use and the self-administered questionnaire is better.

NOTES

Study area: In case one is studying a sensitive topic like harassment at work- a self administered questionnaire is suggested. However, in case the study topic needs additional probing then in that case, a schedule is better.

NOTES

Check Your Progress

1. Why can a questionnaire be used to collect information from a large sample?
2. What is the first and foremost requirement for designing a questionnaire?
3. Define formalized and concealed questionnaire.

5.3 PROCESS OF QUESTIONNAIRE DESIGN

Even though the questionnaire method is most used by researchers, designing a well-structured instrument needs considerable skill. Presented below is a standardized process that a researcher can follow.

Figure 5.1 summarizes the steps involved in questionnaire design.

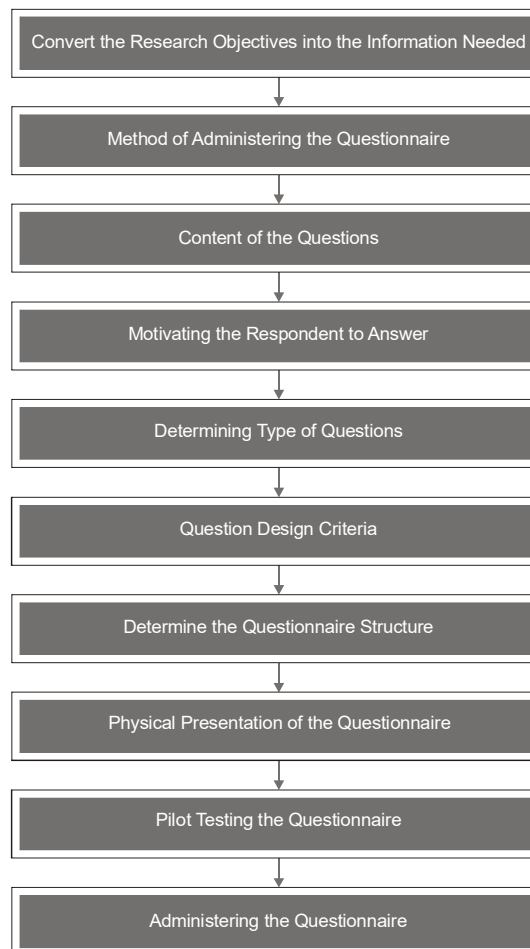


Fig. 5.1 Questionnaire Design Process

1. Information Required Convert the Research objectives into Information Areas

Designing Questionnaire

This is the first step of the design process. By this time the researcher is clear about the research questions; research objectives; variables to be studied; research information required and the characteristics of the population being studied. Once these tasks are done, one can prepare a tabled framework so that the questions which need to be developed become clear. This step-wise process is explained with an example in Table 5.2.

NOTES

Table 5.2 Framework for Identifying Information Needs

Research Questions	Research Objectives	Variables to be Studied	Information (Primary Required)	Population to be Studied
What is the nature of plastic bag usage amongst people in the NCR (National Capital Region)?	To identify the different uses of plastic bags. To find out the method of disposal of plastic bags. To find out who uses plastic bags. To find out what is the level of consciousness that people have about the environment.	Usage behaviour Demographic details	Uses of plastic bags Disposal of plastic bags	Consumers Retailers

2. Method of Administration including Interview Method

Once the researcher has identified his information area; he needs to specify how the information should be collected. The researcher usually has available to him a variety of methods for administering the study. The main methods are personal schedule, self-administered questionnaire through mail, fax, e-mail and web-based questionnaire. There are different preconditions for using one method over the other (Table 5.3).

Table 5.3 Mode of Administration and Design Implications

	Schedule	Telephone	Mail/Fax	E-mail	Web-Based
Administrative control	high	medium	Low	low	low
Sensitive issues	high	medium	Low	low	low
New concept	high	medium	Low	low	low
Large sample	low	low	High	high	high
Cost/time taken	high	medium	Medium	low	low
Question structure	unstructured	either	structured	structured	structured
Sampling control	high	high	Medium	low	low
Response rate	high	high	Low	medium	low
Interviewer bias	high	high	low	low	low

Self-Instructional
Material

As you can observe from Table 5.3, there are many different methods of administering a questionnaire. As mentioned in the previous section, schedule is a type of interview. And as discussed in Unit 4, personal interview is a one-to-one interaction between the investigator/interviewer and the interviewee. The purpose of the dialogue is research specific and ranges from completely unstructured to highly structured. The definition of the structure depends upon the information needs of the research study.

The next step is to determine the matter to be included as questions in the measure. The researcher needs to do an objective quality check in order to see what research objective/information need the question would be covering before using any of the framed questions.

Sometimes, especially in self-administered questionnaires, one may ask some neutral questions at the beginning of the questionnaire to establish an involvement and rapport. For example, for a biofertilizer usage study, the following question was asked:

Do we need to ask several questions instead of a single one? After deciding on the significance of the question, one needs to ascertain whether a single question will serve the purpose or should more than one question be asked. For example, in a TV serial study, one may give ten popular serials to be ranked as 1 to 10 in order of preference. Then the second question after the ranking question is:

88 *Self-Instructional
Material*

Here, one lady might say, 'Everyone in my family watches it'. While another might say, 'It deals with the problems of living in a typical Indian joint family system' and yet another might say, 'My friend recommended it to me'.

Thus, we need to ask her:

'What do you like about _____?'

'Who all in your household watch the serial?'

and

'How did you first hear about the serial?' (Correct)

NOTES

4. Motivating the Respondent to Answer

The questionnaire should be designed in a manner that it involves the respondent and motivates him/her to give information. There are different situations which might lead to this. Each of these is examined separately here:

Does the person have the required information? It has been found that the person has had no experience with the issue being studied. Look at the following question:

How do you evaluate the negotiation skills module, viz., the communication and presentation skill module? (Incorrect)

In this case it might be that the person has not undergone one or even both the modules, so how can he compare? Thus, certain qualifying or filter questions must be asked. Filter questions enable the researcher to filter out the respondents who are not adequately informed. Thus, the correct question would have been:

Have you been through the following training modules?

- Negotiation skills module Yes/no
- Communication and presentation skills Yes/no

In case the answer to both is yes, please answer the following question, or else move to the next question.

How do you evaluate the negotiation skills module, viz., the communication and presentation skill module? (Correct)

Does the person remember? Many a times, the question addressed might be putting too much stress on an individual's memory. For example, consider the following questions:

How much did you spend on eating out last month? (Incorrect)

Such questions are beyond any normal individual's memory bank. Thus, the questions listed above could have been rephrased as follows:

When you go out to eat, on an average your bill amount is:

Less than ₹100

₹101–250

₹251–500

More than ₹500

NOTES

How often do you eat out in a week?

1–2 times

3–4 times

5–6 times

Every day

(correct)

Can the respondent articulate? Sometimes the respondent might not know how to put the answer in clear words. For example, if you ask a respondent to:

- Describe a river rafting experience.

Most respondents would not know what phrases to use to give an answer. Thus, in the above case, one can provide answer categories to the person as follows:

Describe the river rafting experience.

(Correct)

1	Unexciting		Exciting
2	Bad		Good
3	Boring		Interesting
4	Cheap		Expensive
5	Safe		Dangerous

Sensitive information: There might be instances when the question being asked might be embarrassing to the respondents and thus they would not be comfortable in disclosing the data required.

For example, questions such as the following will not get any answers.

Have you ever used fake receipts to claim your medical allowance?
(Incorrect)

Have you ever spit tobacco on the road (to tobacco consumers)?
(Incorrect)

However, in case the socially undesirable habit is in the context of a third person, the chances of getting some correct responses are possible. Thus the questions should be rephrased as follows:

Do you associate with people who use fake receipts to claim their medical allowance? (Correct)

Do you think tobacco consumers spit tobacco on the road? (Correct)

5. Determining the Type of Questions or Questionnaire Composition

Available to the researcher are different kinds of question-response options (Figure 5.2)

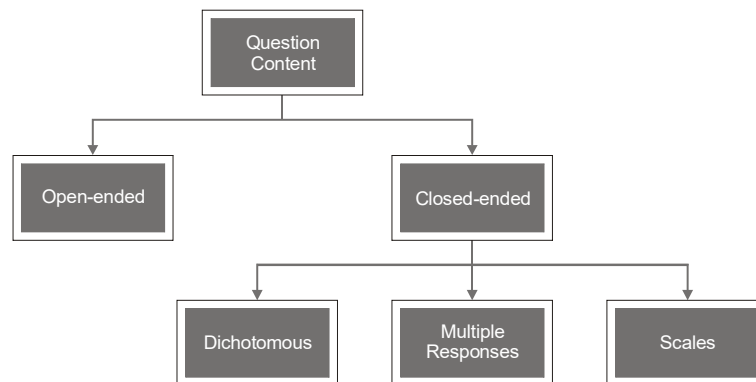


Fig. 5.2 Types of Question–Response Options

NOTES

Open-ended questions

In open-ended questions, the openness refers to the option of answering in one's own words. They are also referred to as unstructured questions or free-response or free-answer questions. Some illustrations of this type are listed below:

- What is your age?
- Which is your favourite TV serial?
- I like Nescafe because _____
- My career goal is to _____

Closed-ended questions

In closed-ended questions, both the question and response formats are structured and defined. There are three kinds of formats as we observed earlier—dichotomous questions, multiple-choice questions and those that have a scaled response.

i. Dichotomous questions: These are restrictive alternatives and provide the respondents only with two answers. These could be 'yes' or 'no', like or dislike, similar or different, married or unmarried, etc.

Are you diabetic?	Yes/No
Have you read the new book by Dan Brown?	Yes/no
What kind of petrol do you use in your car?	Normal/Premium

Dichotomous questions are the easiest type of questions to code and analyze. They are based on the nominal level of measurement and are categorical or binary in nature.

ii. Multiple-choice questions: Unlike dichotomous questions, the person is given a number of response alternatives here. He might be asked to choose the one that is most applicable. For example, this question was given to a retailer who is currently not selling organic food products:

Will you consider selling organic food products in your store?

- Definitely not in the next one year
- Undecided
- Definitely in the next one year
- Probably not in the next one year
- Probably in the next one year

NOTES

Sometimes, multiple-choice questions do not have verbal but rather numerical options for the respondent to choose from, for example:

How much do you spend on grocery products (average in one month)?

Less than ₹2500/-

Between ₹2500–5000/-

More than ₹5000/-

Most multiple-choice questions are based upon ordinal or interval level of measurement. There could also be instances when multiple options are given to the respondent and he can select all those that apply in the case. These kinds of multiple-choice questions are called *checklists*. For example, in the organic food study, the retailer who does not stock organic products was given multiple reasons as follows:

You do not currently sell organic food products because (Could be ≥ 1)

- You do not know about organic food products.
- You are not interested.
- Organic products do not have attractive packaging.
- Organic food products are not supplied regularly.
- Any other _____

iii. Scales: Scales refer to the attitudinal scales that provide a quantitative measurement of attitudes through numerical scores given by researchers to people's responses. The following is a question which has two sub-questions designed on the Likert scale. These require simple agreement and disagreement on the part of the respondent. This scale is based on the interval level of measurement.

Given below are statements related to your organization. Please indicate your agreement/disagreement with each:

(1-Strongly Disagree → → → 5 Strongly Agree)	1	2	3	4	5
1. The people in my company know their roles very clearly.					
2. I want to complete my current task by hook or by crook.					

6. Criteria for Question Designing or Individual Question Content

Step six of the questionnaire involves translating the questions identified into meaningful questions. There are certain designing criteria that a researcher should keep in mind when writing the research questions.

Clearly specify the issue: By reading the question, the person should be able to clearly understand the information need.

Which newspaper do you read? (Incorrect)

This might seem to be a well-defined and structured question. However, the 'you' could be the person filling the questionnaire or the family. He could be reading different newspapers. He might be reading different papers at home and may be the college library. A better way to word the question would be:

Which newspaper or newspapers did you personally read at home during the last month? In case of more than one newspaper, please list all that you read. (Correct)

Use simple terminology: The researcher must take care to ask questions in a language that is understood by the population under study. Technical words or difficult words that are not used in everyday communication must be avoided.

Do you think thermal wear provides immunity? (Incorrect)

Do you think that thermal wear provides you protection from the cold? (Correct)

Avoid ambiguity in questioning: The words used in the questionnaire should mean the same thing to all those answering the questionnaire. A lot of words are subjective and relative in meaning. Consider the following question:

How often do you visit Pizza Hut?

Never

Occasionally

Sometimes

Often

Regularly (Incorrect)

These are ambiguous measures, as *occasionally* in the above question, might be three to four times in a week for one person it, while for another it could be three times in a month. A much better wording for this question would be the following:

In a typical month, how often do you visit Pizza Hut?

Less than once

1 or 2 times

3 or 4 times

More than 4 times (Correct)

Avoid leading questions: Any question that provides a clue to the respondents in terms of the direction in which one wants them to answer is called a leading or biasing question. For example, 'Do you think that working mothers should buy ready-to-eat food when that might contain some chemical preservatives?'

Yes

No

Don't know (Incorrect)

NOTES

NOTES

The question would mostly generate a negative answer, as no working mother would like to buy something that is convenient but might be harmful. Thus, it is advisable to construct a neutral question as follows:

Do you think that working mothers should buy ready-to-eat food?

Yes

No

Don't know

(Correct)

Avoid loaded questions: Questions that address sensitive issues are termed as loaded questions and the response to these questions might not always be honest, as the person might not wish to admit the answer. For example, questions such as follows will rarely get an affirmative answer:

Will you take dowry when you get married?

(Incorrect)

Sensitive questions like this can be rephrased in a variety of ways. For example, the question could be constructed in the context of a third person as follows:

Do you think most Indian men would take dowry when they get married?

(Correct)

Avoid double-barrelled questions: Questions that have two separate options separated by an 'or' or 'and' like the following:

Do you think Nokia and Samsung have a wide variety of touch phones? Yes/No (Incorrect)

The problem is that the respondent might believe that Nokia has better phones or Samsung has better phones or both. These questions are referred to as double-barrelled and the researcher should always split them into two separate questions. For example,

A wide variety of touch phones is available for:

Nokia

Samsung

Both

(Correct)

7. Determine the Questionnaire Structure or Question Order

The questions now have to be put together in a proper sequence.

Instructions: The questionnaires always, even the schedules, begin with standardized instructions. These begin by greeting the respondent and then introducing the researcher and then the purpose of questionnaire administration. For example, in the study on organic food products, the following instructions were given at the beginning of the questionnaire:

'Hi. We _____ are carrying out a market research on the purchase behaviour of grocery products/organic food. We are conducting a survey of consumers, retailers and experts in the NCR for the same.

As you are involved in the purchase and/or consumption of food products, we seek your cooperation for providing the following relevant information for our research. Thank you very much.'

Opening questions: After instructions come the opening questions, which lead the reader into the study topic. For example, a questionnaire on understanding the consumer's buying behavior in malls can ask an opening question that is generic in nature, such as:

What is your opinion about shopping at a mall?

Study questions: After the opening questions, the bulk of the instrument needs to be devoted to the main questions that are related to the specific information needs of the study. Here also, the general rule is that the simpler questions, which do not require a lot of thinking or response time should be asked first as they build the tempo for answering the more difficult/sensitive questions later on. This method of going in a sequential manner from the general to the specific is called the funnel approach.

Classification information: This is the information that is related to the basic socio-economic and demographic traits of the person. These might include name (kept optional in some cases), address, e-mail address and telephone number.

Acknowledgement: The questionnaire ends by acknowledging the inputs of the respondent and thanking him for his cooperation and valuable contribution.

8. Physical Characteristics of the Questionnaire or Form and Layout

The researcher must pay special attention to the look of the questionnaire. The first thing is the quality of the paper on which the questionnaire is printed which should be of good quality. The font style and spacing used in the entire document should be uniform. One must ensure that every question and its response options are printed on the same page. Surveys for different groups could be on different coloured papers. For example, if Delhi is being studied as five zones, then the questionnaire used in each zone could be printed on a differently coloured paper. Each question and section must be numbered properly. In case there is any response instruction for an individual question, it must be before the question. In case the questionnaire is going to be administered by the investigator and if there are any probing questions then they should be clearly written as instructions for the investigator.

9. Pilot Testing of the Questionnaire

Pilot testing refers to testing and administering the designed instrument on a small group of people from the population under study. This is to essentially cover any errors that might have still remained even after the earlier eight steps. For example, the question wording may not be clear, the sequence of questions may not be correct or the question is not needed as it does not solve any purpose. Thus these aspects need to be corrected. Every aspect of the questionnaire has to be tested

NOTES

NOTES

and one must record all the experiences of the conduction, including the time taken to administer it. Sometimes, the researcher might also get the questionnaire whetted by academic or industry experts for their inputs. As far as possible, the pilot should be a small scale replica of the actual survey that would be subsequently conducted.

10. Administering the questionnaire

Once all the nine steps have been completed, the final instrument is ready for conduction and the questionnaire needs to be administered according to the sampling plan.

Check Your Progress

4. What does openness, in open-ended questions, refer to?
5. What are checklists?
6. What kind of language should be used in a questionnaire?
7. Define funnel approach.
8. What does pilot testing refer to?

5.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Since it includes a standard set of questions, it can be successfully used to collect information from a large sample in a reasonably short time period.
2. The first and foremost requirement is that the spelt-out research objectives must be converted into clear questions which will extract answers from the respondent.
3. Formalized and concealed questionnaires have a formal method of questioning; however the purpose is not clear to the respondent. The research studies which are trying to find out the latent causes of behaviour and cannot rely on direct questions use these.
4. In open-ended questions, the openness refers to the option of answering in one's own words. They are also referred to as unstructured questions or free-response or free-answer questions.
5. There could also be instances when multiple options are given to the respondent and he can select all those that apply in the case. These kinds of multiple-choice questions are called *checklists*.
6. The researcher must take care to ask questions in a language that is understood by the population under study. Technical words or difficult words that are not used in everyday communication must be avoided.

7. The method of going in a sequential manner from the general to the specific is called the funnel approach. Here, the simpler questions, which do not require a lot of thinking or response time are asked first as they build the tempo for answering the more difficult/sensitive questions later on.
8. Pilot testing refers to testing and administering the designed instrument on a small group of people from the population under study. This is to essentially cover any errors that might have still remained even after the earlier eight steps.

NOTES

5.5 SUMMARY

- The questionnaire is a research technique that consists of a series of questions asked to respondents, in order to obtain statistically useful information about a given topic.
- However, the use of questionnaire is not always the best method in all research studies. For example, at the exploratory stage, rather than questionnaire, it is advisable to use a more unstructured interview.
- The questions should be self-explanatory and not confusing as then the person will answer the way he understood the question and not in terms of what was asked.
- The two criteria that are most frequently used for designing purposes are the degree of structure and the degree of concealment. Structure refers to the degree to which the response category has been defined. Concealment refers to the degree to which the purpose of the study is explained to the respondent.
- The research studies which are trying to find out the latent causes of behaviour and cannot rely on direct questions use these. For example young people cannot be asked direct questions on whether they are likely to indulge in corruption at work.
- In case the population is illiterate or unable to write the responses, then one must as a rule use the schedule, as the questionnaire cannot be effectively answered by the subject himself.
- The researcher usually has available to him a variety of methods for administering the study. The main methods are personal schedule, self-administered questionnaire through mail, fax, e-mail and web-based questionnaire.
- Sometimes, especially in self-administered questionnaires, one may ask some neutral questions at the beginning of the questionnaire to establish an involvement and rapport.
- There might be instances when the question being asked might be embarrassing to the respondents and thus they would not be comfortable in disclosing the data required.

NOTES

- In closed-ended questions, both the question and response formats are structured and defined. There are three kinds of formats as we observed earlier—dichotomous questions, multiple-choice questions and those that have a scaled response.
- The words used in the questionnaire should mean the same thing to all those answering the questionnaire. A lot of words are subjective and relative in meaning.
- Any question that provides a clue to the respondents in terms of the direction in which one wants them to answer is called a leading or biasing question
- The questionnaires always, even the schedules, begin with standardized instructions. These begin by greeting the respondent and then introducing the researcher and then the purpose of questionnaire administration.
- After the opening questions, the bulk of the instrument needs to be devoted to the main questions that are related to the specific information needs of the study.
- The researcher must pay special attention to the look of the questionnaire. The first thing is the quality of the paper on which the questionnaire is printed which should be of good quality. The font style and spacing used in the entire document should be uniform.
- Pilot testing refers to testing and administering the designed instrument on a small group of people from the population under study. This is to essentially cover any errors that might have still remained even after the earlier eight steps.

5.6 KEY WORDS

- **Questionnaire:** It is a research technique that consists of a series of questions asked to respondents, in order to obtain statistically useful information about a given topic.
- **Concealment:** It refers to the degree to which the purpose of the study is explained to the respondent.
- **Self-administered Questionnaire:** In this questionnaire, the respondent reads all the instructions and questions on his own and records his own statements or responses. It is designed specifically to be completed by a respondent without intervention of the researchers (e.g. an interviewer) collecting the data.
- **Attitude Scales:** These are scales that provide a quantitative measurement of attitudes, opinions or values by summarising numerical scores given by researchers to people's responses to sets of statements exploring dimensions of an underlying theme.

5.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write a short note on the two criteria used for designing purposes.
2. When is a non-formalized and concealed questionnaire used?
3. Write a short note on dichotomous questions.
4. What are loaded questions?
5. Define classification information.

Long-Answer Questions

1. Discuss the criteria that must be kept in mind while designing the questionnaire.
2. Describe the prerequisites on which the choice of a schedule or self-administered questionnaire depends.
3. Explain the first three steps involved in questionnaire design.
4. Analyse the structure of a questionnaire.

5.8 FURTHER READINGS

- Saravanavel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.
- Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.
- Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.
- Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.
- Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

NOTES

NOTES

UNIT 6 MEASUREMENT AND SCALING TECHNIQUES

Structure

- 6.0 Introduction
- 6.1 Objectives
- 6.2 Scales of Measurement
 - 6.2.1 Importance of Measurement and Scaling in Marketing Research
- 6.3 Primary Scales of Measurement and their Fundamental Properties
- 6.4 Rating and Attitude Measurement Scales
- 6.5 Types of Comparative Scales and Non-Comparative Scales
- 6.6 Selecting an Appropriate Scale and Scale Evaluation
- 6.7 Answers to Check Your Progress Questions
- 6.8 Summary
- 6.9 Key Words
- 6.10 Self Assessment Questions and Exercises
- 6.11 Further Readings

6.0 INTRODUCTION

The term ‘measurement’ is concerned with assigning numbers to an observation for the purpose of providing useful and accurate information. It plays an important role in marketing research by measuring the perceptions, attitude and preferences of customers towards a particular product. These further help in developing responses to marketing related problems such as price change, product design, brand loyalty, product quality and so on. Some scales of measurement used for this purpose are nominal scale, ordinal scale, interval scale and ratio scale. This unit focuses on the different types of scales of measurement and the ways of selecting an appropriate scale.

6.1 OBJECTIVES

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

- Discuss the Scales of Measurement
- Analyse the primary Scales of Measurement and their fundamental properties
- Explain the types of Comparative and Non-Comparative Scales
- Describe the process of selecting an appropriate Scale and Scale Evaluation

6.2 SCALES OF MEASUREMENT

The term ‘measurement’ means assigning numbers or some other symbols to the characteristics of certain objects. When numbers are used, the researcher must have a rule for assigning a number to an observation in a way that provides an accurate description. We do not measure the object but some characteristics of it. Therefore, in research, people/consumers are not measured; what is measured are only their perceptions, attitude or any other relevant characteristics. There are two reasons for which numbers are usually assigned. First of all, numbers permit statistical analysis of the resulting data and secondly, they facilitate the communication of measurement results.

As mentioned earlier, the numbering is done based on certain rules. Therefore, the assignment of numbers to the characteristics must be isomorphic, i.e., there must be a one-to-one correspondence between the numbers and the characteristics being measured.

For example, same rupee figures should be assigned to a household with identical annual income. Only then numbers can be associated with specific characteristics of the measured object and vice versa. Further, they must not change over the objects or time. This means that the rules for a given assignment must be invariant over time or the object being measured.

Scaling is an extension of measurement. Scaling involves creating a continuum on which measurements on objects are located. Suppose you want to measure the satisfaction level towards Kingfisher Airlines and a scale of 1 to 11 is used for the said purpose. This scale indicates the degree of dissatisfaction, with 1 = extremely dissatisfied and 11 = extremely satisfied. Measurement is the actual assignment of a number from 1 to 11 to each respondent whereas the scaling is the process of placing the respondent on a continuum with respect to their satisfaction towards Kingfisher Airlines.

Attitude

An attitude is viewed as an enduring disposition to respond consistently in a given manner to various aspects of the world, including persons, events and objects. A company is able to sell its products or services when its customers have a favourable attitude towards its products/services. In the reverse scenario, the company will not be able to sustain itself for long. It, therefore, becomes very important to measure the attitude of the customers towards the company’s products/services. Unfortunately, attitude cannot be measured directly. In order to measure an attitude, we make an inference based on the perceptions the customers have about the product/services. The attitude is derived from the perceptions. If the consumers have a favourable perception towards the products/services, the attitude will be favourable. Therefore, the attitudes are indirectly observed.

NOTES

Basically, attitude has three components: cognitive, affective and intention (or action) components.

6.2.1 Importance of Measurement and Scaling in Marketing Research

NOTES

In eliciting opinion, concentration may be paid to the current or contemporary topics. An increased attention is being paid to the attitude surveys by the marketing researchers, they have been found to be extremely useful in the matters of various marketing problems such as response to an advertisement, price change, product quality, product design, brand loyalty, trade mark of both durable and non-durable goods. As more and more tendency towards objective analysis of marketing problems develops, such surveys grow widely, and would form the basis of testing existing hypothesis and thereon enunciate new ones.

Merits of Attitude Measurement

1. A controlled and planned marketing development *needs the knowledge of attitude of consumers*. In other words, the knowledge of response mechanism is extremely essential for successful implementation of the marketing plans.
2. The attitude of consumers would aid in successful *classification of 'types' of consumers*. The basis of grouping is the attitude of consumers towards a certain product or service. This implies their needs with quantitateness and saves a lot of effort as well as labour.
3. *In framing a questionnaire or schedule for conducting any survey successfully, one needs the knowledge on consumer's attitude*. The researcher on knowing the weak points of the respondents only can know when they are likely to give false reply, under what situations, how can their cooperation be achieved, and so on. A thorough and systematic study of attitudes would aid the researcher to guard against pit falls.
4. *Attitude affords predictability* which is very much useful to control marketing conditions.
5. Peaceful and standard living are the aims of marketing, which need to know the attitude of consumers as well as producers, their emotional reactions and behaviour pattern in different situations. What is true of individuals is true for groups also.
6. The study of attitude has a practical importance. *A good salesman must be well versed with attitude of consumers*, and understands how different types of consumers behave. A knowledge of this and timely use may turn for achieving good results.

Limitations of Attitude Measurement

1. *The attitude is intangible and not subject to visual observation*. What a consumer feels and how he will react to given product or service is only a guess work.

2. *The consumer attitude is a complex affair* due to multiple influences. Hence, we cannot say with certainty how a person will react.
3. *Measuring attitude lacks proper scale.* Like physical sciences, marketing research has no instrument devised infallibly to measure attitude correctly.
4. *Difficulty also arises due to the changing nature of attitude* which is caused due to a multiplicity of factors influencing the attitude. The attitude of same person may vary with variation in time and circumstances.

NOTES

Check Your Progress

1. What are the reasons for which numbers are assigned in a research?
2. What does a controlled and planned marketing development need?

6.3 PRIMARY SCALES OF MEASUREMENT AND THEIR FUNDAMENTAL PROPERTIES

There are four types of measurement scales—nominal, ordinal, interval and ratio. These are called primary scales of measurement. We will discuss each one of them in detail. The choice of the measurement scale has implications for the statistical technique to be used for data analysis.

Nominal scale: This is the lowest level of measurement. Here, numbers are assigned for the purpose of identification of the objects. Any object which is assigned a higher number is in no way superior to the one which is assigned a lower number. Each number is assigned to only one object and each object has only one number assigned to it. It may be noted that the objects are divided into mutually exclusive and collectively exhaustive categories.

Example:

- What is your religion?
 - (a) Hinduism
 - (b) Sikhism
 - (c) Christianity
 - (d) Islam
 - (e) Any other, (please specify)

A Hindu may be assigned a number 1, a Sikh may be assigned a number 2, a Christian may be assigned a number 3 and so on. Any religion which is assigned a higher number is in no way superior to the one which is assigned a lower number. The assignment of numbers is only for the purpose of identification.

Nominal scale measurements are used for identifying food habits (vegetarian or non-vegetarian), gender (male/female), caste, respondents, marital status, brands, attributes, stores, the players of a hockey team and so on.

NOTES

The assigned numbers cannot be added, subtracted, multiplied or divided. The only arithmetic operations that can be carried out are the count of each category. Therefore, a frequency distribution table can be prepared for the nominal scale variables and mode of the distribution can be worked out. One can also use chi-square test and compute contingency coefficient using nominal scale variables.

Ordinal scale: This is the next higher level of measurement than the nominal scale measurement. One of the limitations of the nominal scale measurements is that we cannot say whether the assigned number to an object is higher or lower than the one assigned to another option. The ordinal scale measurement takes care of this limitation. An ordinal scale measurement tells whether an object has more or less of characteristics than some other objects. However, it cannot answer how much more or how much less.

Example:

- Rank the following attributes while choosing a restaurant for dinner. The most important attribute may be ranked one, the next important may be assigned a rank of 2 and so on.

Attribute	Rank
Food quality	
Prices	
Menu variety	
Ambience	
Service	

In the ordinal scale, the assigned ranks cannot be added, multiplied, subtracted or divided. One can compute median, percentiles and quartiles of the distribution. The other major statistical analyses which can be carried out is the rank order correlation coefficient, sign test. All the statistical techniques which are applicable in the case of nominal scale measurement can also be used for the ordinal scale measurement. However, the reverse is not true. This is because ordinal scale data can be converted into nominal scale data but not the other way round.

Interval scale: The interval scale measurement is the next higher level of measurement. It takes care of the limitation of the ordinal scale measurement where the difference between the score on the ordinal scale does not have any meaningful interpretation. In the interval scale the difference of the score on the scale has meaningful interpretation. It is assumed that the respondent is able to answer the questions on a continuum scale. The mathematical form of the data on the interval scale may be written as

$$Y = a + bX \quad \text{Where } a \neq 0$$

In the interval scale, the difference in score has a meaningful interpretation while the ratio of the score on this scale does not have a meaningful interpretation. This can be seen from the following interval scale question:

- How likely are you to buy a new designer carpet in the next six months?

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Scale A	1	2	3	4	5
Scale B	0	1	2	3	4
Scale C	-2	-1	0	1	2

Suppose a respondent ticks the response category 'likely' and another respondent ticks the category 'unlikely'. If we use any of the scales A, B or C, we note that the difference between the scores in each case is 2. Whereas, when the ratio of the scores is taken, it is 2, 3 and -1 for the scales A, B and C respectively. Therefore, the ratio of the scores on the scale does not have a meaningful interpretation. The following are some examples of interval scale data.

- How important is price to you while buying a car?

Least important	Unimportant	Neutral	Important	Most important
1	2	3	4	5

- How do you rate the work environment of your organization?

Very good	Good	Neither good nor bad	Bad	Very bad
5	4	3	2	1

- How expensive is the restaurant 'Punjabi By Nature'?

Extremely expensive	Definitely expensive	Somewhat expensive	Somewhat inexpensive	Definitely inexpensive	Extremely inexpensive
1	2	3	4	5	6

The numbers on this scale can be added, subtracted, multiplied or divided. One can compute arithmetic mean, standard deviation, correlation coefficient and conduct a t-test, Z-test, regression analysis and factor analysis. As the interval scale data can be converted into the ordinal and the nominal scale data, therefore all the techniques applicable for the ordinal and the nominal scale data can also be used for interval scale data.

Ratio scale: This is the highest level of measurement and takes care of the limitations of the interval scale measurement, where the ratio of the measurements on the scale does not have a meaningful interpretation. The ratio scale measurement can be converted into interval, ordinal and nominal scale. But the other way round is not possible. The mathematical form of the ratio scale data is given by $Y = bX$. In this case, there is a natural zero (origin), whereas in the interval scale we had an

NOTES

NOTES

arbitrary zero. Examples of the ratio scale data are weight, distance travelled, income and sales of a company, to mention a few.

All the mathematical operations can be carried out using the ratio scale data. In addition to the statistical analysis mentioned in the interval, the ordinal and the nominal scale data, one can compute coefficient of variation, geometric mean and harmonic mean using the ratio scale measurement.

Check Your Progress

3. What is the limitation of nominal scale measurement?
4. State the analyses which can be carried out in ordinal scale.
5. Which limitation of interval scale measurement is taken care of by ratio scale?

6.4 RATING AND ATTITUDE MEASUREMENT SCALES

Apart from the primary scales, some commonly used scaling methods fall under the categories of Rating measurement scales and Attitude measurement scales. In this section, this categorization is briefly discussed and will be explained in detail in the next section.

Rating Scale

In a rating scale, the rater evaluates a person or object or both phenomenon on point along a continuum or in a category. A numeric, value is then assigned to this point or category.

- (a) **Graphic rating scale** Raters mark or indicate how they feel on a graphic scale of some sort. A common graphic scale is known as the thermometer chart.
- (b) **Itemized rating scale** Raters select one of a limited number of categories that are ordered in some fashion. The number of categories usually varies somewhat between two and eleven points.
- (c) **Comparative rating scale** Raters judge a person, object or other phenomenon against some present standard or some other person, object or other phenomenon. This scale can take a variety of forms. One of them is called a *rank-order* scale.
- (d) **Paired comparison rating scale** The rater makes a number of comparisons in which he selects between objects presented to him two at a time.

- (e) **Constant-sum rating scale** Here, the rates are required to allocate a given number of objects according to some criterion. The technique is best used for relatively small number of objects.
- (f) **Fractionation rating scale** It requires the raters to rate an object by comparing it with a reference point. The goal is a ratio-scale representation of attitudes toward the complete set of objects.

Attitude scales are those that measure an individual's predisposition toward any person, object or other phenomenon. These scales differ from rating scales (even though rating is an intricate part of attitude scaling) in that they are generally more complex, multi-item scales.

- (a) **Likert scale:** Respondent indicates degree of agreement or disagreement with a variety of statements about some attitude, object, person, or event. Usually the scales are five or seven points. The scales are summed across statements to get the attitude score.
- (b) **Semantic differential scale:** Respondent rates an attitude object on a number of five or seven point bipolar adjectives or phrases. The selection of adjectives or phrases is based on the attitude object, person or event.
- (c) **Stapel scale:** It is a modification of the semantic differential scale. It differs from the semantic differential scale in that the adjectives of descriptive phrases are tested separately instead of simultaneously as bipolar pairs. Also, the points on the scale are identified by number and there are ten scale positions rather than five or seven as in the earlier case.
- (d) **Thurstone scale:** Thurstone developed his method of equal-appearing intervals on the concept that even though people could not assign quantitative measurements to their own attitudes, they could tell the difference between the attitude represented by two different statements and could identify items that were approximately halfway between the two. But these scales are not widely used in marketing research, probably because of the time-consuming task of preparing the scale statements. Also it does not obtain information about the degree or intensity of agreement with the various statements.
- (e) **Scalogram analysis:** It is another method for scaling attitude items. Here, the attitude items can be arranged in such an order that a respondent who positively answers to a particular question also responds positively to all other items having a lower rank. It is based on the assumption that the ordering of certain stimuli is possible.

NOTES

Check Your Progress

6. How does the Stapel scale differ from semantic differential scale?
7. Why are Thurstone scales not widely used in marketing research?

6.5 TYPES OF COMPARATIVE SCALES AND NON-COMPARATIVE SCALES

NOTES

The scaling techniques used in research can also be classified into comparative and non-comparative scales (Figure 6.1).

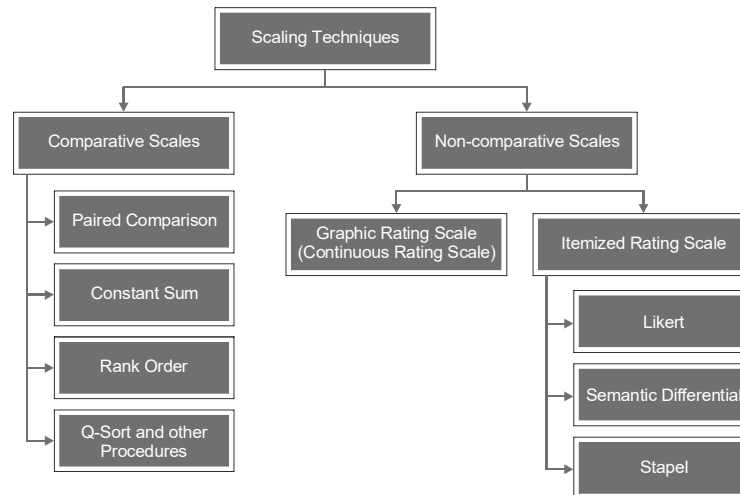


Fig. 6.1 Types of Scaling Techniques

I. Comparative Scales

In comparative scales, it is assumed that respondents make use of a standard frame of reference before answering the question. For example:

A question like ‘How do you rate Barista in comparison to Cafe Coffee Day on quality of beverages?’ is an example of the comparative rating scale. It involves the direct comparison of stimulus objects. Example:

- Please rate Domino’s in comparison to Pizza Hut on the basis of your satisfaction level on an 11-point scale, based on the following parameters: (1 = Extremely poor, 6 = Average, 11 = Extremely good). Circle your response:

a.	Variety of menu options	1	2	3	4	5	6	7	8	9	10	11
b.	Value for money	1	2	3	4	5	6	7	8	9	10	11
c.	Speed of service (delivery time)	1	2	3	4	5	6	7	8	9	10	11
d.	Promotional offers	1	2	3	4	5	6	7	8	9	10	11
e.	Food quality	1	2	3	4	5	6	7	8	9	10	11

Comparative scale data is interpreted generally in a relative kind. Below are discussed each of the scales under comparative rating scales in detail below. Bear in mind, you have already been introduced to some of there under the rating and attitude measurement scale classification in the previous section.

(i) Paired comparison scales: Here a respondent is presented with two objects and is asked to select one according to whatever criterion he or she wants to use. The resulting data from this scale is ordinal in nature. As an example, suppose a

parent wants to offer one of the four items to a child—chocolate, burger, ice cream and pizza. The child is offered to choose one out of the two from the six possible pairs, i.e., chocolate or burger, chocolate or ice cream, chocolate or pizza, burger or ice cream, burger or pizza and ice cream or pizza. In general, if there are n items, the number of paired comparison would be $(n(n-1)/2)$. Paired comparison technique is useful when the number of items is limited because it requires a direct comparison and overt choice.

(ii) Rank order scaling: In the rank order scaling, respondents are presented with several objects simultaneously and asked to order or rank them according to some criterion. Consider, for example the following question:

- Rank the following soft drinks in order of your preference, the most preferred soft drink should be ranked one, the second most preferred should be ranked two and so on.

Soft Drinks	Rank
Coke	
Pepsi	
Limca	
Sprite	
Mirinda	
Seven Up	
Fanta	

Like paired comparison, this approach is also comparative in nature. The problem with this scale is that if a respondent does not like any of the above-mentioned soft drink and is forced to rank them in the order of his choice, then, the soft drink which is ranked one should be treated as the least disliked soft drink and similarly, the other rankings can be interpreted. The rank order scaling results in the ordinal data.

(iii) Constant sum rating scaling: In constant sum rating scale, the respondents are asked to allocate a total of 100 points between various objects and brands. The respondent distributes the points to the various objects in the order of his preference. Consider the following example:

- Allocate a total of 100 points among the various schools into which you would like to admit your child. The points should be allocated in such a way that the sum total of the points allocated to various schools adds up to 100.

Schools	Points
DPS	
Mother's International	
APEEJAY	
DAV Public School	
Laxman Public School	
TOTAL POINTS	100

NOTES

NOTES

Suppose Mother's International is awarded 30 points, whereas Laxman Public School is awarded 15 points, one can make a statement that the respondent rates Mother's International twice as high as Laxman Public School. This type of data is not only comparative in nature but could also result in ratio scale measurement.

(iv) Q-sort technique: This technique makes use of the rank order procedure in which objects are sorted into different piles based on their similarity with respect to certain criterion. Suppose there are 100 statements and an individual is asked to pile them into five groups, in such a way, that the strongly agreed statements could be put in one pile, agreed statements could be put in another pile, neutral statements form the third pile, disagreed statements come in the fourth pile and strongly disagreed statements form the fifth pile, and so on. The data generated in this way would be ordinal in nature. The distribution of the number of statements in each pile should be such that the resulting data may follow a normal distribution.

II. Non-Comparative Scales

In the non-comparative scales, the respondents do not make use of any frame of reference before answering the questions. The resulting data is generally assumed to be interval or ratio scale.

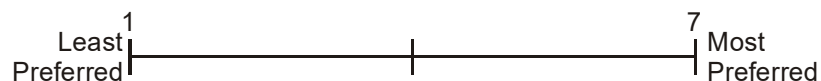
The non-comparative scales are divided into two categories, namely, the graphic rating scales and the itemized rating scales. A useful and widely used itemized rating scale is the Likert scale.

Let's discuss each of them. Bear in mind, you have already learnt about them briefly in the previous section.

(i) Graphic Rating Scale

This is a continuous scale, also called graphic rating Scale. In the graphic rating scale, the respondent is asked to tick his preference on a graph. Consider for example the following question:

- Please put a tick mark (✓) on the following line to indicate your preference for fast food.



To measure the preference of an individual towards the fast food one has to measure the distance from the extreme left to the position where a tick mark has been put. Higher the distance, higher would be the individual preference for fast food. This scale suffers from two limitations—one, if a respondent has put a tick mark at a particular position and after ten minutes, he or she is given another form to put a tick mark, it will virtually be impossible to put a tick at the same position as was done earlier. Does it mean that the respondent's preference for fast food has undergone a change in 10 minutes? The basic assumption in this scale is that the respondents can distinguish the fine shade in differences between the preference/attitude which need not be the case. Further, the coding, editing and tabulation of

data generated through such a procedure is a very tedious task and researchers try to avoid using it.

(ii) Itemized Rating Scale

In the itemized rating scale, the respondents are provided with a scale that has a number of brief descriptions associated with each of the response categories. The response categories are ordered in terms of the scale position and the respondents are supposed to select the specified category that describes in the best possible way an object is rated. There are certain issues that should be kept in mind while designing the itemized rating scale. These issues are:

Number of categories to be used: There is no hard and fast rule as to how many categories should be used in an itemized rating scale. However, it is a practice to use five or six categories. Some researches are of the opinion that more than five categories should be used in situations where small changes in attitudes are to be measured. There are others that argue that the respondents would find it difficult to distinguish between more than five categories.

Odd or even number of categories: It has been a matter of debate among the researchers as to whether an odd or even number of categories are to be used. By using even number of categories, the scale would not have a neutral category and the respondent will be forced to choose either the positive or the negative side of the attitude. If odd numbers of categories are used, the respondent has the freedom to be neutral if he wants to be so.

Balanced versus unbalanced scales: A balanced scale is the one which has an equal number of favourable and unfavourable categories. The following is the example of a balanced scale:

- How important is price to you in buying a new car?

Very important
Relatively important
Neither important nor unimportant
Relatively unimportant
Very unimportant

In this question, there are five response categories, two of which emphasize the importance of price and two others that do not show its importance. The middle category is neutral.

The following is the example of the unbalanced scale.

- How important is price to you in buying a new car?

More important than any other factor
Extremely important
Important
Somewhat important
Unimportant

NOTES

NOTES

In this question there are four response categories that are skewed towards the importance given to the price, whereas one category is for the unimportant side. Therefore, this question is an unbalanced question.

Nature and degree of verbal description: Many researchers believe that each category must have a verbal, numerical or pictorial description. Verbal description should be clearly and precisely worded so that the respondents are able to differentiate between them. Further, the researcher must decide whether to label every scale category, some scale categories, or only extreme scale categories.

Forced versus non-forced scales: In the forced scale, the respondent is forced to take a stand, whereas in the non-forced scale, the respondent can be neutral if he/she so desires. The argument for a forced scale is that those who are reluctant to reveal their attitude are encouraged to do so with the forced scale. Paired comparison scale, rank order scale and constant sum rating scales are examples of forced scales.

Physical form: There are many options that are available for the presentation of the scales. It could be presented vertically or horizontally. The categories could be expressed in boxes, discrete lines or as units on a continuum. They may or may not have numbers assigned to them. The numerical values, if used, may be positive, negative or both.

Suppose we want to measure the perception about Jet Airways using a multi-item scale. One of the questions is about the behaviour of the crew members. Given below is a set of scale configurations that may be used to measure their behaviour.

The behaviour of the crew members of Jet Airways is:

- | | | | | | | | |
|----|--------------------------|----------------------|----------------------|-------|-------|-----------|-----------|
| 1. | Very bad | _____ | _____ | _____ | _____ | _____ | Very good |
| 2. | Very bad | 1 | 2 | 3 | 4 | 5 | Very good |
| 3. | ☐ | Very bad | | | | | |
| | ☐ | | | | | | |
| | ☐ | Neither bad nor good | | | | | |
| | ☐ | | | | | | |
| | ☐ | Very good | | | | | |
| 4. | Very bad | Bad | Neither bad nor good | | Good | Very Good | |
| 5. | -2 | -1 | 0 | | 1 | 2 | |
| | Very bad | | Neither bad nor good | | | Very good | |

Check Your Progress

8. What does Q-sort technique make use of?
9. State the categories of non-comparative scales.
10. What is the argument for a forced scale?

6.6 SELECTING AN APPROPRIATE SCALE AND SCALE EVALUATION

The selection of the appropriate scale depends on the following factors:

- The Objective of the research
- The properties of the data
- The nature of attitudinal object to be measured
- Manager's problem formulation
- Whether rating, sorting, ranking or choice is required?
- Whether labels are involved?
- Whether forced or non-forced scales are to be used?
- Whether balanced or unbalanced are required?

There are three criteria for evaluating measurements: reliability, validity and sensitivity. It may be noted that there is a relationship between reliability and sensitivity. If we want to make an item more sensitive, it may be achieved at the cost of reliability. This means to get more sensitivity, the researcher might have to compromise with reliability.

1. Reliability

Reliability is concerned with consistency, accuracy and predictability of the scale. It refers to the extent to which a measurement process is free from random errors. The reliability of a scale can be measured using the following methods:

Test-retest reliability: In this method, repeated measurements of the same person or group using the same scale under similar conditions are taken. A very high correlation between the two scores indicates that the scale is reliable. The researcher has to be careful in deciding the time difference between two observations. If the time difference between two observations is very small, it is very likely that the respondent would give same answer which could result in higher correlation. Further if the difference is too large, the attitude might have undergone a change during that period, resulting in a weak correlation and hence poor reliability. Therefore, researchers have to be very careful in deciding the time difference between observation. Generally, a time difference of about 5-6 months is considered as an ideal period.

Split-half reliability method: This method is used in the case of multiple item scales. Here the number of items is randomly divided into two parts and a correlation coefficient between the two is obtained. A high correlation indicates that the internal consistency of the construct leads to greater reliability.

2. Validity

The validity of a scale refers to the question: whether we are measuring what we want to measure. Validity of the scale refers to the extent to which the measurement process is free from both systematic and random errors. The validity of a scale is

NOTES

NOTES

a more serious issue than reliability. There are different ways to measure validity.

Content validity: This is also called face validity. It involves subjective judgement by an expert for assessing the appropriateness of the construct. For example, to measure the perception of a customer towards Kingfisher Airlines, a multiple item scale is developed. A set of 15 items is proposed. These items when combined in an index measure the perception of Kingfisher Airlines. In order to judge the content validity of these 15 items, a set of experts may be requested to examine the representativeness of the 15 items. The items covered may be lacking in the content validity if we have omitted behaviour of the crew, food quality, and food quantity, etc., from the list. In fact, conducting the exploratory research to exhaust the list of items measuring perception of the airline would be of immense help in such a case.

Predictive validity: This involves the ability of a measured phenomenon at one point of time to predict another phenomenon at a future point of time. If the correlation coefficient between the two is high, the initial measure is said to have a high predictive ability. As an example, consider the use of the common admission test (CAT) to shortlist candidates for admission to the MBA programme in a business school. The CAT scores are supposed to predict the candidate's aptitude for studies towards business education.

3. Sensitivity

Sensitivity refers to an instrument's ability to accurately measure the variability in a concept. A dichotomous response category such as agree or disagree does not allow the recording of any attitude changes. A more sensitive measure with numerous categories on the scale may be required. For example, adding 'strongly agree', 'agree', 'neither agree nor disagree', 'disagree' and 'strongly disagree' categories will increase the sensitivity of the scale.

The sensitivity of scale based on a single question or a single item can be increased by adding questions or items. In other words, because composite measures allow for a greater range of possible scores, they are more sensitive than a single-item scale.

Check Your Progress

11. State the three criteria for evaluating measurements.
12. What does validity of a scale refer to?
13. What does predictive validity involve?

6.7 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. There are two reasons for which numbers are usually assigned. First of all, numbers permit statistical analysis of the resulting data and secondly, they facilitate the communication of measurement results.

2. A controlled and planned marketing development *needs the knowledge of attitude of consumers*. In other words, the knowledge of response mechanism is essential for successful implementation of the marketing plans.
3. The limitation of the nominal scale measurement is that we cannot say whether the assigned number to an object is higher or lower than the one assigned to another option.
4. One can compute median, percentiles and quartiles of the distribution in ordinal scale. The other major statistical analyses which can be carried out is the rank order correlation coefficient, sign test.
5. Ratio scale is the highest level of measurement and takes care of the limitations of the interval scale measurement, where the ratio of the measurements on the scale does not have a meaningful interpretation.
6. Stapel scale is a modification of the semantic differential scale. It differs from the semantic differential scale in that the adjectives of descriptive phrases are tested separately instead of simultaneously as bipolar pairs. Also, the points on the scale are identified by number and there are ten scale positions rather than five or seven as in the earlier case.
7. Thurstone scales are not widely used in marketing research, probably because of the time-consuming task of preparing the scale statements. Also it does not obtain information about the degree or intensity of agreement with the various statements.
8. The Q-sort technique makes use of the rank order procedure in which objects are sorted into different piles based on their similarity with respect to certain criterion.
9. The non-comparative scales are divided into two categories, namely, the graphic rating scales and the itemized rating scales.
10. The argument for a forced scale is that those who are reluctant to reveal their attitude are encouraged to do so with the forced scale.
11. There are three criteria for evaluating measurements: reliability, validity and sensitivity.
12. The validity of a scale refers to the question: whether we are measuring what we want to measure. Validity of the scale refers to the extent to which the measurement process is free from both systematic and random errors.
13. Predictive validity involves the ability of a measured phenomenon at one point of time to predict another phenomenon at a future point of time.

NOTES

6.8 SUMMARY

- Numbering is done based on certain rules. Therefore, the assignment of numbers to the characteristics must be isomorphic, i.e., there must be a

NOTES

- one-to-one correspondence between the numbers and the characteristics being measured.
- Scaling is an extension of measurement. Scaling involves creating a continuum on which measurements on objects are located.
 - In framing questionnaire or schedule for conducting any survey successfully needs knowledge on consumer's attitude.
 - The study of attitude has a practical importance. A good salesman must be well versed with attitude of consumers, and understands how different types of consumers behave.
 - Nominal scale is the lowest level of measurement. Here, numbers are assigned for the purpose of identification of the objects. Any object which is assigned a higher number is in no way superior to the one which is assigned a lower number.
 - An ordinal scale measurement tells whether an object has more or less of characteristics than some other objects. However, it cannot answer how much more or how much less.
 - In the interval scale, the difference in score has a meaningful interpretation while the ratio of the score on this scale does not have a meaningful interpretation.
 - Ratio scale is the highest level of measurement and takes care of the limitations of the interval scale measurement, where the ratio of the measurements on the scale does not have a meaningful interpretation.
 - An attitude is viewed as an enduring disposition to respond consistently in a given manner to various aspects of the world, including persons, events and objects. A company is able to sell its products or services when its customers have a favourable attitude towards its products/services.
 - Cognitive component represents an individual's information and knowledge about an object. It includes awareness of the existence of the object, beliefs about the characteristics or attributes of the object and judgement about the relative importance of each of the attributes.
 - Attitude scales are those that measure an individual's predisposition toward any person, object or other phenomenon. These scales differ from rating scales (even though rating is an intricate part of attitude scaling) in that they are generally more complex, multi-item scales.
 - In scalogram analysis, the attitude items can be arranged in such an order that a respondent who positively answers to a particular question also responds positively to all other items having a lower rank.
 - In comparative scales, it is assumed that respondents make use of a standard frame of reference before answering the question.

- In constant sum rating scale, the respondents are asked to allocate a total of 100 points between various objects and brands.
- The non-comparative scales are divided into two categories, namely, the graphic rating scales and the itemized rating scales. A useful and widely used itemized rating scale is the Likert scale.
- In the forced scale, the respondent is forced to take a stand, whereas in the non-forced scale, the respondent can be neutral if he/she so desires.
- Reliability is concerned with consistency, accuracy and predictability of the scale. It refers to the extent to which a measurement process is free from random errors.
- The validity of a scale refers to the question whether we are measuring what we want to measure. Validity of the scale refers to the extent to which the measurement process is free from both systematic and random errors. The validity of a scale is a more serious issue than reliability. There are different ways to measure validity.

NOTES

6.9 KEY WORDS

- **Likert Scale:** It is a unidimensional scale that researchers use to collect respondents' attitudes and opinions. Researchers often use this psychometric scale to understand the views and perspectives towards a brand, product, or target market.
- **Scalogram Analysis:** It is an ordinal scale type where statements are arranged in a hierarchical order so that someone who agrees with one item will also agree with lower-order, easier, less extreme items.
- **Unaided recall:** It is a marketing technique to determine how well a consumer remembers an advertisement without any external help such as clues, or visuals.
- **Aided recall:** It is a question-asking strategy in which survey respondents are provided with a number of cues to facilitate their memory of particular responses that are of relevance to the purpose of the study.
- **Balanced Scale:** It is a scale which has an equal number of favourable and unfavourable categories.

6.10 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. List two limitations of attitude measurement.
2. What are nominal scale measurements used for?

NOTES

3. What is done in order to measure an attitude?
4. What does the intention component of attitude reflect?
5. On the basis of which concept did Thurstone develop the Thurstone scale?
6. How are reliability and sensitivity related?

Long-Answer Questions

1. Discuss the merits of attitude measurement.
2. Explain the three components of attitude.
3. Describe the issues to be kept in mind while designing the itemized rating scale.
4. Elaborate upon the test-retest reliability method.

6.11 FURTHER READINGS

Saravanel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.

Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.

Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.

Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.

Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

UNIT 7 SAMPLING THEORY: DESIGN, SIZE AND TECHNIQUES

*Sampling Theory:
Design, Size and
Techniques*

NOTES

Structure

- 7.0 Introduction
- 7.1 Objectives
- 7.2 Meaning of Sampling
 - 7.2.1 Importance of Sampling in Marketing Research
- 7.3 Basic Constructs of Sampling
- 7.4 Determining the Sample Size
- 7.5 Process of Sampling Design: Probability and Non-Probability Sampling and their Classification
- 7.6 Selecting an Appropriate Sampling Technique
- 7.7 Answers to Check Your Progress Questions
- 7.8 Summary
- 7.9 Key Words
- 7.10 Self Assessment Questions and Exercises
- 7.11 Further Readings

7.0 INTRODUCTION

In planning an enquiry, the researcher has to determine as to whether the investigation is to take account of the whole population or only a part of it. A survey may be conducted by either of the two methods: Census Method or Sampling Methods. The census method implies that the investigator investigates all the cases in the universe. Sampling on the other hand only considers a sample from a given universe.

Sampling theory aims to provide the most effective ways of acquiring a sample that will give the most useful and accurate results regarding the surveyed population. Sampling forms an important part of marketing research because it provides an idea about the opinions of a large group by surveying a small sample among them. This information helps a firm to align its marketing decisions as per the scenario in the chosen market. The basic constructs of sampling are: population, element, sampling frame, sample, sampling unit, sampling and census. This unit provides an analysis of the meaning, importance and basic constructs of sampling. Moreover, the process of sampling design and the selection of an appropriate sampling design have also been described.

NOTES

7.1 OBJECTIVES

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

- Explain meaning and basic constructs of Sampling
- Analyze the methods for determining the Sample Size
- Discuss the process of Sampling Design
- Describe the ways to select an appropriate Sampling Technique

7.2 MEANING OF SAMPLING

Sampling refers to the method of selecting a sample from a given universe with a view to draw conclusions about the universe. *Sample* means representative of universe selected for the study. The sampling is a key of research and has acquired the status of a technical job. Values in population are known are *parameters*; those in samples are termed *statistics*.

Sampling is based on the law of statistical regularity and the law of inertia of large number. Under the first law, if a small number of items (called a *sample*) are picked up from a large number of items (the larger group called the *universe or population*), the sample will tend to possess almost the same characteristics as the large groups are more stable than small groups because we have the compensating effect of deviations in opposable directions. Of course, our sample should be large enough to represent truly the entire population or the universe.

A sample is a selection of units from the entire group called the population or universe of interest. In marketing, a sample is a particular segment or part of the market, and it is the focus of the marketing decisions which may be applied to the entire market. The units included in the sample may be individual respondents, housewives, families, dealers, firms, city blocks, and so on. The units included in the sample may be easily approached to obtain the desired information with a view to taking marketing decisions.

A sampling plan is a vital element of a research design. It must indicate:

- (i) *The sampling unit*, i.e. who is to be our respondent or who is to be surveyed;
- (ii) *The sample size*, i.e. how many respondents are to be contacted under the survey;
- (iii) *The sampling procedures*, i.e. how they are to be selected; and
- (iv) *The sampling media*, i.e. how they are to be approached with a view to obtaining the desired information.

The person who is most likely to have the desired information will be our respondent or unit in the sample. This person may be the housewife, the husband,

the family or the retailer. The sample size must be sufficiently large so that we may have a representative sample.

Sampling is widely used in market research and its methodologies were pioneered many years ago by social investigators such as Bowley who undertook notable enquiries into the conditions of the working class in reading. This disciplined approach strongly influenced the methodology of social surveys, which became more systematic in the use of sampling technologies. Sampling theory is based on the relationships between a population (not necessarily, as stated already, a human population, the term is used in its statistical sense) and the samples drawn from it. Probability theory allows certain conclusions to be drawn about a specific population and the samples taken from it.

The marketing researcher must have clear understanding of the following: (i) defining universe, (ii) sampling unit, (iii) source list, (iv) sampling size, (v) sampling methods, and (vi) sampling errors.

7.2.1 Importance of Sampling in Marketing Research

The following are the benefits which make sampling important to marketing research:

- **More reliable results:** The sampling technique ensures more reliable result than obtained from a complete count. Indeed, even a complete census can only be tested for accuracy by sampling check.
- **More detailed information:** The sampling technique makes sure to collect more detailed and reliable information covering a wide range of universe.
- **Save time and labour:** This method is a part of the population; a considerable time and labour are saved. Time saved not only in collecting data but also is processing of it.
- **Save cost:** The amount of effort and expense involved in collecting information is always lesser for unit of the sample than a complete census i.e., the total financial burden of sample survey is generally less than that of census method.
- **Only method:** There are some instances in which the census method is inapplicable, and the only practicable means is provided by the sample method.

Marketing researchers consider sampling important as it helps them to function efficiently at the global scale by learning about the mass customer base in different countries without investing a lot of time, money and labour. In today's market scenario, information is considered key and gaining maximum relevant information about the attitudes and preferences of wide customer base is possible through sampling methods. Additionally, market researchers use the sampling methods to gain vital information about the product and its use in the market, in what manner they can be modified or upgraded depending on customer attitudes in different markets as well as gain information to develop new products.

NOTES

NOTES

Check Your Progress

1. Name the laws on which sampling is based.
2. What is a sample in marketing?
3. List the points about which a marketing researcher must have a clear understanding.

7.3 BASIC CONSTRUCTS OF SAMPLING

The process of selecting the right individuals, objects or events for a study is known as sampling. Sampling involves the study of a small number of individuals, objects chosen from a larger group. Before we get into the details of various issues pertaining to sampling, it would be appropriate to discuss some of the basic constructs of sampling.

- **Population:** Population refers to any group of people or objects that form the subject of study in a particular survey and are similar in one or more ways. For example, the number of full-time MBA students in a business school could form one population. If there are 200 such students, the population size would be 200. We may be interested in understanding their perceptions about business education. If, in an organization, there are 1,000 engineers, out of which 350 are mechanical engineers and we are interested in examining the proportion of mechanical engineers who intend to leave the organization within six months, all the 350 mechanical engineers would form the population of interest. If the interest is in studying how the patients in a hospital are looked after, then all the patients of the hospital would fall under the category of population.
- **Element:** An element comprises a single member of the population. Out of the 350 mechanical engineers mentioned above, each mechanical engineer would form an element of the population.
- **Sampling frame:** Sampling frame comprises all the elements of a population with proper identification that is available to us for selection at any stage of sampling. For example, the list of registered voters in a constituency could form a sampling frame; the telephone directory; the number of students registered with a university; the attendance sheet of a particular class and the payroll of an organization are examples of sampling frames. When the population size is very large, it becomes virtually impossible to form a sampling frame. We know that the number of consumers of soft drinks is very large and, therefore, it becomes very difficult to form the sampling frame for the same.
- **Sample:** It is a subset of the population. It comprises only some elements of the population. If out of the 350 mechanical engineers employed in an

organization, 30 are surveyed regarding their intention to leave the organization in the next six months, these 30 members would constitute the sample.

- **Sampling unit:** A sampling unit is a single member of the sample. If a sample of 50 students is taken from a population of 200 MBA students in a business school, then each of the 50 students is a sampling unit.
- **Sampling:** It is a process of selecting an adequate number of elements from the population so that the study of the sample will not only help in understanding the characteristics of the population but also enable us to generalize the results. We will see later that there are two types of sampling designs—probability sampling design and non-probability sampling design.
- **Census (or complete enumeration):** An examination of each and every element of the population is called census or complete enumeration. Census is an alternative to sampling. We will discuss the inherent advantages of sampling over a complete enumeration later.

NOTES

Sample vs Census

In a research study, we are generally interested in studying the characteristics of a population. Suppose there are 2 lakh households in a town, and we are interested in estimating the proportion of households that spend their summer vacations on a hill station. This information can be obtained by asking every household in that town. If all the households in a population are asked to provide information, such a survey is called a census. There is an alternative way of obtaining the same information, by choosing a subset of all the two lakh households and asking them for the same information. This subset is called a sample. Based upon the information obtained from the sample, a generalization about the population characteristics could be made. However, that sample has to be representative of the population. For a sample to be representative of the population, the distribution of sampling units in the sample has to be in the same proportion as the elements in the population. For example, if in a town there are 50, 35 and 15 per cent households in lower, middle and upper income groups respectively, then a sample taken from this population should have the same proportions for it to be representative. As mentioned earlier, there are several advantages of a sample over census.

- Sample saves time and cost. Many times a decision-maker may not have too much of time to wait till all the information is available. Therefore, a sample could come to his rescue.
- There are situations where a sample is the only option. When we want to estimate the average life of fluorescent bulbs, what is done is that they are burnt out completely. If we go for a complete enumeration there would not be anything left for use. Another example could be testing the quality of a photographic film.

NOTES

- The study of a sample instead of complete enumeration may, at times, produce more reliable results. This is because by studying a sample, fatigue is reduced and fewer errors occur while collecting the data, especially when a large number of elements are involved.

A census is appropriate when the population size is small, e.g., the number of public sector banks in the country. Suppose the researcher is interested in collecting information from the top management of a bank regarding their views on the monetary policy announced by the Reserve Bank of India (RBI), in this case, a complete enumeration may be possible as the population size is not very large.

Sampling vs Non-Sampling Error

There are two types of errors that may occur while we are trying to estimate the population parameters from the sample. These are called sampling and non-sampling errors.

Sampling error: This error arises when a sample is not representative of the population. It is the difference between sample mean and population mean. The sampling error reduces with the increase in sample size as an increased sample may result in increasing the representativeness of the sample.

Non-sampling error: This error arises not because a sample is not a representative of the population but because of other reasons. Some of these reasons are listed below:

- The respondents when asked for information on a particular variable may not give the correct answers. If a person aged 48 is asked a question about his age, he may indicate the age to be 36, which may result in an error in estimating the true value of the variable of interest.
- The error can arise while transferring the data from the questionnaire to the spreadsheet on the computer.
- There can be errors at the time of coding, tabulation and computation.
- If the population of the study is not properly defined, it could lead to errors.
- The chosen respondent may not be available to answer the questions or may refuse to be part of the study.

Check Your Progress

4. Why does the study of a sample, instead of complete enumeration, produce more reliable results?
5. When does a sampling error arise?
6. List the reasons due to which a non-sampling error arises.

7.4 DETERMINING THE SAMPLE SIZE

Sampling Theory:
Design, Size and
Techniques

The size of a sample depends upon the basic characteristics of the population, the type of information required from the survey and the cost involved. Therefore, a sample may vary in size for several reasons. The size of the population does not influence the size of the sample as will be shown later on.

NOTES

There are various methods of determining the sample size in practice:

- Researchers may arbitrarily decide the size of sample without giving any explicit consideration to the accuracy of the sample results or the cost of sampling. This arbitrary approach should be avoided.
- For some of the projects, the total budget for the field survey (usually mentioned) in a project proposal is allocated. If the cost of sampling per sample unit is known, one can easily obtain the sample size by dividing the total budget allocation by the cost of sampling per unit. This method concentrates only on the cost aspect of sampling, rather than the value of information obtained from such a sample.
- There are other researchers who decide on the sample size based on what was done by the other researchers in similar studies. Again, this approach cannot be a substitute for the formal scientific approach.
- The most commonly used approach for determining the size of sample is the confidence interval approach covered under inferential statistics. Below will be discussed this approach while determining the size of a sample for estimating population mean and population proportion. In a confidence interval approach, the following points are taken into account for determining the sample size in estimation of problems involving means:
 - (a) *The variability of the population:* It would be seen that the higher the variability as measured by the population standard deviation, larger will be the size of the sample. If the standard deviation of the population is unknown, a researcher may use the estimates of the standard deviation from previous studies. Alternatively, the estimates of the population standard deviation can be computed from the sample data.
 - (b) *The confidence attached to the estimate:* It is a matter of judgement, how much confidence you want to attach to your estimate. Assuming a normal distribution, the higher the confidence the researcher wants for the estimate, the larger the sample size. This is because the value of the standard normal ordinate ' Z ' will vary accordingly. For a 90 per cent confidence, the value of ' Z ' would be 1.645 and for a 95 per cent confidence, the corresponding ' Z ' value would be 1.96 and so on (see Table 7.1). It would be seen later that a higher confidence would lead to a larger ' Z ' value.

NOTES

(c) *The allowable error or margin of error*: How accurate do we want our estimate to be is again a matter of judgement of the researcher. It will of course depend upon the objectives of the study and the consequence resulting from the higher inaccuracy. If the researcher seeks greater precision, the resulting sample size would be large.

Table 7.1 Area under standard normal distribution between the mean and successive value of Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4804	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Sample Size for Estimating Population Mean

Sampling Theory:
Design, Size and
Techniques

The formula for determining the sample size in such a case is given by

$$n = \frac{Z^2 \sigma^2}{e^2}$$

Where $\bar{X} - \mu = e$ = Margin of error

n = Sample size

σ = Population standard deviation

Z = the value of standard normal ordinate

It may be noted from above that the size of the sample is directly proportional to the variability in the population and the value of Z for a confidence interval. It varies inversely with the size of the error. It may also be noted that the size of a sample does not depend upon the size of population.

Below is given a worked out example for the determination of a sample size.

Example 7.1: An economist is interested in estimating the average monthly household expenditure on food items by the households of a town. Based on past data, it is estimated that the standard deviation of the population on the monthly expenditure on food item is ₹30. With allowable error set at ₹7, estimate the sample size required at a 90 per cent confidence.

Solution:

90 per cent confidence $\Rightarrow Z = 1.645$

$e = ₹7$

$\sigma = ₹30$

$$\begin{aligned} n &= \frac{Z^2 \sigma^2}{e^2} \\ &= \frac{(1.645)^2 (30)^2}{(7)^2} \\ &= 49.7025 \\ &= 50 \text{ (approx.)} \end{aligned}$$

Determination of Sample Size for Estimating the Population Proportion

The formula for determining the sample size in such a case is given by

$$n = \frac{Z^2 pq}{e^2}$$

The above formula will be used if the value of population proportion (proportion of occurrence of the event) p is known. If, however, p is unknown,

NOTES

we substitute the maximum value of pq in the above formula. It can be shown that the maximum value of pq is $1/4$ when $p = 1/2$ and $q = 1/2$.

NOTES

Therefore,
$$n = \frac{1}{4} \frac{Z^2}{e^2}$$

Let us consider two examples for determining a sample size while estimating the population proportion.

Example 7.2: A manager of a department store would like to study women's spending per year on cosmetics. He is interested in knowing the population proportion of women who purchase their cosmetics primarily from his store. If he wants to have a 90 per cent confidence of estimating the true proportion to be within ± 0.045 , what sample size is needed?

Solution:

$$\begin{aligned} 90 \text{ per cent confidence} &\Rightarrow Z = 1.645 \\ e &= \pm .045 \\ n &= \frac{1}{4} \frac{Z^2}{e^2} \\ &= \frac{1 (1.645)^2}{4 (0.45)^2} \\ &= 334.0772 \\ &= 335 \text{ (approx.)} \end{aligned}$$

Example 7.3: A consumer electronics company wants to determine the job satisfaction levels of its employees. For this, they ask a simple question, 'Are you satisfied with your job?' It was estimated that no more than 30 per cent of the employees would answer yes. What should be the sample size for this company to estimate the population proportion to ensure a 95 per cent confidence in result, and to be within 0.04 of the true population proportion?

Solution:

$$\begin{aligned} 95 \text{ per cent confidence} &\Rightarrow Z = 1.96 \\ e &= 0.04 \\ p &= 0.3 \\ q &= 0.7 \\ n &= \frac{Z^2 pq}{e^2} \\ &= \frac{(1.96)^2 \times 0.3 \times 0.7}{(0.04)^2} \\ &= 504.21 \\ &= 505 \text{ (approx.)} \end{aligned}$$

Points to be noted for sample size determination

There are certain issues to be kept in mind before applying the formulas for the determination of sample size in this unit. First of all, these formulas are applicable for simple random sampling only. Further, they relate to the sample size needed for the estimation of a particular characteristic of interest. In a survey, a researcher needs to estimate several characteristics of interests and each one of them may require a different sample size. In case the universe is divided into different strata, the accuracy required for determining the sample size for each strata may be different. However, the present method will not be able to serve the requirement. Lastly, the formulas for sample size must be based upon adequate information about the universe.

Check Your Progress

7. What does the size of a sample depend upon?
8. What can one do if the cost of sampling per unit sample is known?

NOTES

7.5 PROCESS OF SAMPLING DESIGN: PROBABILITY AND NON-PROBABILITY SAMPLING AND THEIR CLASSIFICATION

Sampling design refers to the process of selecting samples from a population. There are two types of sampling designs—probability sampling design and non-probability sampling design. Probability sampling designs are used in conclusive research. In a probability sampling design, each and every element of the population has a known chance of being selected in the sample. The known chance does not mean equal chance. Simple random sampling is a special case of probability sampling design where every element of the population has both known and equal chance of being selected in the sample.

In case of non-probability sampling design, the elements of the population do not have any known chance of being selected in the sample. These sampling designs are used in exploratory research.

(i) Probability Sampling Design

Under this, the following sampling designs would be covered—simple random sampling with replacement (SRSWR), simple random sampling without replacement (SRSWOR), systematic sampling and stratified random sampling.

(a) Simple random sampling with replacement (SRSWR)

Under this scheme, a list of all the elements of the population from where the samples are to be drawn is prepared. If there are 1,000 elements in the population,

NOTES

we write the identification number or the name of all the 1,000 elements on 1,000 different slips. These are put in a box and shuffled properly. If there are 20 elements to be selected from the population, the simple random sampling procedure involves selecting a slip from the box and reading of the identification number. Once this is done, the chosen slip is put back to the box and again a slip is picked up and the identification number is read from that slip. This process continues till a sample of 20 is selected. Please note that the first element is chosen with a probability of $1/1,000$. The second one is also selected with the same probability and so are all the subsequent elements of the population.

(b) Simple random sampling without replacement (SRSWOR)

In case of simple random sample without replacement, the procedure is identical to what was explained in the case of simple random sampling with replacement. The only difference here is that the chosen slip is not placed back in the box. This way, the first unit would be selected with the probability of $1/1,000$, second unit with the probability of $1/999$, the third will be selected with a probability of $1/998$ and so on, till we select the required number of elements (in this case, 20) in our sample.

The simple random sampling (with or without replacement) is not used in consumer research. This is because in a consumer research the population size is usually very large, which creates problems in the preparation of a sampling frame. For example, number of consumers of soft drinks, pizza, shampoo, soap, chocolate, etc, is very large. However, these (SRSWR and SRSWOR) designs could be useful when the population size is very small, for example, the number of steel/aluminum-producing companies in India and the number of banks in India. Since the population size is quite small, the preparation of a sampling frame does not create any problem.

Another problem with these (SRSWR and SRSWOR) designs is that we may not get a representative sample using such a scheme. Consider an example of a locality having 10,000 households, out of which 5,000 belong to low-income group, 3,500 belong to middle income group and the remaining 1,500 belong to high-income group. Suppose it is decided to take a sample of 100 households using the simple random sampling. The selected sample may not contain even a single household belonging to the high- and middle-income group and only the low-income households may get selected, thus, resulting in a non-representative sample.

(c) Systematic Sampling

Systematic sampling takes care of the limitation of the simple random sampling that the sample may not be a representative one. In this design, the entire population is arranged in a particular order. The order could be the calendar dates or the elements of a population arranged in an ascending or a descending order of the magnitude which may be assumed as random. List of subjects arranged in the alphabetical order could also be used and they are usually assumed to be random

in order. Once this is done, the steps followed in the systematic sampling design are as follows:

- First of all, a sampling interval given by $K = N/n$ is calculated, where N = the size of the population and n = the size of the sample. It is seen that the sampling interval K should be an integer. If it is not, it is rounded off to make it an integer.
- A random number is selected from 1 to K . Let us call it C .
- The first element to be selected from the ordered population would be C , the next element would be $C + K$ and the subsequent one would be $C + 2K$ and so on till a sample of size n is selected.

This way we can get representation from all the classes in the population and overcome the limitations of the simple random sampling. To take an example, assume that there are 1,000 grocery shops in a small town. These shops could be arranged in an ascending order of their sales, with the first shop having the smallest sales and the last shop having the highest sales. If it is decided to take a sample of 50 shops, then our sampling interval K will be equal to $1000 \div 50 = 20$. Now we select a random number from 1 to 20. Suppose the chosen number is 10. This means that the shop number 10 will be selected first and then shop number $10 + 20 = 30$ and the next one would be $10 + (2 \times 20) = 50$ and so on till all the 50 shops are selected. This way we can get a representative sample in the sense that it will contain small, medium and large shops.

It may be noted that in a systematic sampling the first unit of the sample is selected at random (probability sampling design) and having chosen this, we have no control over the subsequent units of sample (non-probability sampling). Because of this, this design at times is called mixed sampling.

The main advantage of systematic sampling design is its simplicity. When sampling from a list of population arranged in a particular order, one can easily choose a random start as described earlier. After having chosen a random start, every K_{th} item can be selected instead of going for a simple random selection. This design is statistically more efficient than a simple random sampling, provided the condition of ordering of the population is satisfied.

The use of systematic sampling is quite common as it is easy and cheap to select a systematic sample. In systematic sampling, one does not have to jump back and forth all over the sampling frame wherever random number leads and neither does one have to check for duplication of elements as compared to simple random sampling. Another advantage of a systematic sampling over simple random sampling is that one does not require a complete sampling frame to draw a systematic sample. The investigator may be instructed to interview every 10th customer entering a mall without a list of all customers.

NOTES

NOTES

(d) Stratified Random Sampling

Under this sampling design, the entire population (universe) is divided into strata (groups), which are mutually exclusive and collectively exhaustive. By mutually exclusive, it is meant that if an element belongs to one stratum, it cannot belong to any other stratum. Strata are collectively exhaustive if all the elements of various strata put together completely cover all the elements of the population. The elements are selected using a simple random sampling independently from each group.

There are two reasons for using a stratified random sampling rather than simple random sampling. One is that the researchers are often interested in obtaining data about the component parts of a universe. For example, the researcher may be interested in knowing the average monthly sales of cell phones in 'large', 'medium' and 'small' stores. In such a case, separate sampling from within each stratum would be called for. The second reason for using a stratified random sampling is that it is more efficient as compared to a simple random sampling. This is because dividing the population into various strata increases the representativeness of the sampling as the elements of each stratum are homogeneous to each other.

There are certain issues that may be of interest while setting up a stratified random sample. These are:

- *What criteria should be used for stratifying the universe (population)?*

The criteria for stratification should be related to the objectives of the study. The entire population should be stratified in such a way that the elements are homogeneous within the strata, whereas there should be heterogeneity between strata. As an example, if the interest is to estimate the expenditure of households on entertainment, the appropriate criteria for stratification would be the household income. This is because the expenditure on entertainment and household income are highly correlated.

Generally, stratification is done on the basis of demographic variables like age, income, education and gender. Customers are usually stratified on the basis of life stages and income levels to study their buying patterns. Companies may be stratified according to size, industry, profits for analysing the stock market reactions.

- *How many strata should be constructed?*

Going by common sense, as many strata as possible should be used so that the elements of each stratum will be as homogeneous as possible. However, it may not be practical to increase the number of strata and, therefore, the number may have to be limited. Too many strata may complicate the survey and make preparation and tabulation difficult. Costs of adding more strata may be more than the benefit obtained. Further, the researcher may end up with the practical difficulty of

preparing a separate sampling frame as the simple random samples are to be drawn from each stratum.

- *What should be appropriate number of samples size to be taken in each stratum?*

This question pertains to the number of observations to be taken out from each stratum. At the outset, one needs to determine the total sample size for the universe and then allocate it between each stratum. This may be explained as follows:

Let there be a population of size N . Let this population be divided into three strata based on a certain criterion. Let N_1 , N_2 and N_3 denote the size of strata 1, 2 and 3 respectively, such that $N = N_1 + N_2 + N_3$. These strata are mutually exclusive and collectively exhaustive. Each of these three strata could be treated as three populations. Now, if a total sample of size n is to be taken from the population, the question arises that how much of the sample should be taken from strata 1, 2 and 3 respectively, so that the sum total of sample sizes from each strata adds up to n .

Let the size of the sample from first, second and third strata be n_1 , n_2 , and n_3 respectively such that $n = n_1 + n_2 + n_3$. Then, there are two schemes that may be used to determine the values of n_i , ($i = 1, 2, 3$) from each strata. These are proportionate and disproportionate allocation schemes.

Proportionate allocation scheme: In this scheme, the size of the sample in each stratum is proportional to the size of the population of the strata. For example, if a bank wants to conduct a survey to understand the problems that its customers are facing, it may be appropriate to divide them into three strata based upon the size of their deposits with the bank. If we have 10,000 customers of a bank in such a way that 1,500 of them are big account holders (having deposits of more than ₹10 lakh), 3,500 of them are medium-sized account holders (having deposits of more than ₹2 lakh but less than ₹10 lakh), the remaining 5,000 are small account holders (having deposits of less than 2 lakh). Suppose the total budget for sampling is fixed at ₹20,000 and the cost of sampling a unit (customer) is ₹20. If a sample of 100 is to be chosen from all the three strata, the size of the sample from strata 1 would be:

$$n_1 = n \times \frac{N_1}{N} = 100 \times \frac{1500}{10000} = 15$$

The size of sample from strata 2 would be:

$$n_2 = n \times \frac{N_2}{N} = 100 \times \frac{3500}{10000} = 35$$

The size of sample from strata 3 would be:

$$n_3 = n \times \frac{N_3}{N} = 100 \times \frac{5000}{10000} = 50$$

NOTES

NOTES

This way the size of the sample chosen from each stratum is proportional to the size of the stratum. Once we have determined the sample size from each stratum, one may use the simple random sampling or the systematic sampling or any other sampling design to take out samples from each of the strata.

Disproportionate allocation: As per the proportionate allocation explained above, the sizes of the samples from strata 1, 2 and 3 are 15, 35 and 50 respectively. As it is known that the cost of sampling of a unit is ₹20 irrespective of the strata from where the sample is drawn, the bank would naturally be more interested in drawing a large sample from stratum 1, which has the big customers, as it gets most of its business from strata 1. In other words, the bank may follow a disproportionate allocation of sample as the importance of each stratum is not the same from the point of view of the bank. The bank may like to take a sample of 45 from strata 1 and 40 and 15 from strata 2 and 3 respectively. Also, a large sample may be desired from the strata having more variability.

(ii) Non-probability Sampling Designs

Under the non-probability sampling, the following designs would be considered—convenience sampling, purposive (judgemental) sampling and snowball sampling.

(a) Convenience Sampling

Convenience sampling is used to obtain information quickly and inexpensively. The only criterion for selecting sampling units in this scheme is the convenience of the researcher or the investigator. Mostly, the convenience samples used are neighbours, friends, family members, colleagues and ‘passers-by’. This sampling design is often used in the pre-test phase of a research study such as the pre-testing of a questionnaire. Some of the examples of convenience sampling are:

- People interviewed in a shopping centre for their political opinion for a TV programme.
- Monitoring the price level in a grocery shop with the objective of inferring the trends in inflation in the economy.
- Requesting people to volunteer to test products.
- Using students or employees of an organization for conducting an experiment.

In all the above situations, the sampling unit may either be self-selected or selected because of ease of availability. No effort is made to choose a representative sample. Therefore, in this design, the difference between the population value (parameters) of interest and the sample value (statistic) is unknown both in terms of the magnitude and direction. Therefore, it is not possible to make an estimate of the sampling error and researchers would not be able to make a conclusive statement about the results from such a sample. It is because of this, convenience sampling should not be used in conclusive research (descriptive and causal research).

Convenience sampling is commonly used in exploratory research. This is because the purpose of an exploratory research is to gain an insight into the problem

and generate a set of hypotheses which could be tested with the help of a conclusive research. When very little is known about a subject, a small-scale convenience sampling can be of use in the exploratory work to help understand the range of variability of responses in a subject area.

(b) Judgemental Sampling

Under judgemental sampling, experts in a particular field choose what they believe to be the best sample for the study in question. The judgement sampling calls for special efforts to locate and gain access to the individuals who have the required information. Here, the judgement of an expert is used to identify a representative sample. For example, the shoppers at a shopping centre may serve to represent the residents of a city or some of the cities may be selected to represent a country. Judgemental sampling design is used when the required information is possessed by a limited number/category of people. This approach may not empirically produce satisfactory results and, may, therefore, curtail generalizability of the findings due to the fact that we are using a sample of experts (respondents) that are usually conveniently available to us. Further, there is no objective way to evaluate the precision of the results. A company wanting to launch a new product may use judgemental sampling for selecting 'experts' who have prior knowledge or experience of similar products. A focus group of such experts may be conducted to get valuable insights. Opinion leaders who are knowledgeable are included in the organizational context. Enlightened opinions (views and knowledge) constitute a rich data source. A very special effort is needed to locate and have access to individuals who possess the required information.

The most common application of judgemental sampling is in business-to-business (B to B) marketing. Here, a very small sample of lead users, key accounts or technologically sophisticated firms or individuals is regularly used to test new product concepts, producing programmes, etc.

(c) Quota Sampling

In quota sampling, the sample includes a minimum number from each specified subgroup in the population. The sample is selected on the basis of certain demographic characteristics such as age, gender, occupation, education, income, etc. The investigator is asked to choose a sample that conforms to these parameters. Field workers are assigned quotas of the sample to be selected satisfying these characteristics.

(d) Snowball Sampling

Snowball sampling is generally used when it is difficult to identify the members of the desired population, e.g., deep-sea divers, families with triplets, people using walking sticks, doctors specializing in a particular ailment, etc. Under this design, each respondent, after being interviewed, is asked to identify one or more in the field. This could result in a very useful sample. The main problem is in making the

NOTES

NOTES

initial contact. Once this is done, these cases identify more members of the population, who then identify further members and so on. It may be difficult to get a representative sample. One plausible reason for this could be that the initial respondents may identify other potential respondents who are similar to themselves. The next problem is to identify new cases.

Check Your Progress

9. Which designs are covered under probability sampling design?
10. Why is simple random sampling not used in consumer research?
11. What are the two reasons for using a stratified random sampling rather than simple random sampling?
12. Why is convenience sampling commonly used in exploratory research?

7.6 SELECTING AN APPROPRIATE SAMPLING TECHNIQUE

In the foregoing pages, we discussed some major types of sample designs, highlighting their distinguishing features. The reader may like to know what are the characteristics of a good sample design. A good sample design requires the judicious balancing of four broad criteria : goal orientation, measurability, practicality and economy.

1. **Goal Orientation:** This suggests that a sample design “should be oriented to the research objectives, tailored to the survey design, and fitted to the survey conditions”. If this is done, it should influence the choice of the population, the measurement as also the procedure of choosing a sample.
2. **Measurability:** A sample design should enable the computation of valid estimates of its sampling variability. Normally, this variability is expressed in the form of standard errors in surveys. However, this is possible only in the case of probability sampling. In non-probability samples, such as a quota sample, it is not possible to know the degree of precision of the survey results.
3. **Practicality:** This implies that the sample design can be followed properly in the survey, as envisaged earlier. It is necessary that complete, correct, practical and clear instructions should be given to the interviewer so that no mistakes are made in the selection of sampling units and the final selection in the field is not different from the original sample design. Practicality also refers to simplicity of the design, i.e., it should be capable of being understood and followed in actual operation of the field work.
4. **Economy:** Finally, economy implies that the objectives of the survey should be achieved with minimum cost and effort. Survey objectives are generally

spelt out in terms of precision, i.e., the inverse of the variance of survey estimates. For a given degree of precision, the sample design should give the minimum cost. Alternatively, for a given per unit cost, the sample design should achieve maximum precision (minimum variance).

It may be pointed out that these four criteria come into conflict with each other in most of the cases, and the researcher should carefully balance the conflicting criteria so that he is able to select a really good sample design. As there is no unique method or procedure by which one can select a good sample, one has to compare several sample designs that can be used in a survey. This means that one has to weigh the pros and cons, the strong and weak points of various sample designs in respect of these four criteria, before selecting the best possible one.

Check Your Progress

13. What does a good sample design require?
14. What does goal orientation suggest?

7.7 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Sampling is based on the law of statistical regularity and the law of inertia of large number.
2. In marketing, a sample is a particular segment or part of the market, and it is the focus of the marketing decisions which may be applied to the entire market.
3. The marketing researcher must have clear understanding of the following:
(i) defining universe, (ii) sampling unit, (iii) source list, (iv) sampling size, (v) sampling methods, and (vi) sampling errors.
4. The study of a sample instead of complete enumeration may, at times, produce more reliable results because by studying a sample, fatigue is reduced and fewer errors occur while collecting the data, especially when a large number of elements are involved.
5. Sampling error arises when a sample is not representative of the population.
6. The reasons due to which a non-sampling error arises are:
 - The error can arise while transferring the data from the questionnaire to the spreadsheet on the computer.
 - There can be errors at the time of coding, tabulation and computation.
7. The size of a sample depends upon the basic characteristics of the population, the type of information required from the survey and the cost involved.

NOTES

NOTES

8. If the cost of sampling per sample unit is known, one can easily obtain the sample size by dividing the total budget allocation by the cost of sampling per unit.
9. Under probability sampling design, the following sampling designs would be covered—simple random sampling with replacement (SRSWR), simple random sampling without replacement (SRSWOR), systematic sampling and stratified random sampling.
10. The simple random sampling (with or without replacement) is not used in consumer research because in a consumer research, the population size is usually very large, which creates problems in the preparation of a sampling frame.
11. The two reasons for using a stratified random sampling rather than simple random sampling are: One is that the researchers are often interested in obtaining data about the component parts of a universe. The second reason for using a stratified random sampling is that it is more efficient as compared to a simple random sampling.
12. Convenience sampling is commonly used in exploratory research because the purpose of an exploratory research is to gain an insight into the problem and generate a set of hypotheses which could be tested with the help of a conclusive research.
13. A good sample design requires the judicious balancing of four broad criteria: goal orientation, measurability, practicality and economy.
14. Goal orientation suggests that a sample design “should be oriented to the research objectives, tailored to the survey design, and fitted to the survey conditions”.

7.8 SUMMARY

- Sampling refers to the method of selecting a sample from a given universe with a view to draw conclusions about the universe. *Sample* means representative of universe selected for the study.
- In marketing, a sample is a particular segment or part of the market, and it is the focus of the marketing decisions which may be applied to the entire market.
- The unit of sample – a house, a family, an individual or a group, must be decided before drawing sample. The units should be clear, capable of being objectively ascertained.
- The process of selecting the right individuals, objects or events for a study is known as sampling. Sampling involves the study of a small number of individuals, objects chosen from a larger group.

- Sample saves time and cost. Many times a decision-maker may not have too much of time to wait till all the information is available. Therefore, a sample could come to his rescue.
- A census is appropriate when the population size is small, e.g., the number of public sector banks in the country.
- Sampling error arises when a sample is not representative of the population. It is the difference between sample mean and population mean. The sampling error reduces with the increase in sample size as an increased sample may result in increasing the representativeness of the sample.
- The size of a sample depends upon the basic characteristics of the population, the type of information required from the survey and the cost involved. Therefore, a sample may vary in size for several reasons. The size of the population does not influence the size of the sample.
- Sampling design refers to the process of selecting samples from a population. There are two types of sampling designs—probability sampling design and non-probability sampling design. Probability sampling designs are used in conclusive research.
- In case of non-probability sampling design, the elements of the population do not have any known chance of being selected in the sample. These sampling designs are used in exploratory research.
- Systematic sampling takes care of the limitation of the simple random sampling that the sample may not be a representative one. In this design, the entire population is arranged in a particular order.
- The main advantage of systematic sampling design is its simplicity. When sampling from a list of population arranged in a particular order, one can easily choose a random start as described earlier.
- The criteria for stratification should be related to the objectives of the study. The entire population should be stratified in such a way that the elements are homogeneous within the strata, whereas there should be heterogeneity between strata.
- In proportionate allocation scheme, the size of the sample in each stratum is proportional to the size of the population of the strata.
- Convenience sampling is used to obtain information quickly and inexpensively. The only criterion for selecting sampling units in this scheme is the convenience of the researcher or the investigator.
- Under judgemental sampling, experts in a particular field choose what they believe to be the best sample for the study in question. The judgement sampling calls for special efforts to locate and gain access to the individuals who have the required information.

NOTES

NOTES

- In quota sampling, the sample includes a minimum number from each specified subgroup in the population. The sample is selected on the basis of certain demographic characteristics such as age, gender, occupation, education, income, etc.
- A sample design should enable the computation of valid estimates of its sampling variability. Normally, this variability is expressed in the form of standard errors in surveys.

7.9 KEY WORDS

- **Sampling:** It refers to the method of selecting a sample from a given universe with a view to draw conclusions about the universe.
- **Source List:** It is a list which contains the names of the units of universe from which the sample is to be selected.
- **Proportionate Allocation:** It is a scheme in which the size of the sample in each stratum is proportional to the size of the population of the strata.
- **Judgement Sampling:** It is a type of random sample that is selected based on the opinion of an expert. Results obtained from a judgment sample are subject to some degree of bias, due to the frame and population not being identical

7.10 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write a short note on the law of statistical regularity.
2. What should a sampling plan indicate?
3. What is the most commonly used approach for determining the size of sample?
4. Why is systematic sampling called mixed sampling?
5. Why should convenience sampling not be used in conclusive research?

Long-Answer Questions

1. Discuss some concepts of sampling
2. Explain the advantages of a sample over census.
3. Analyse the issues that may be of interest while setting up a stratified random sample.
4. Elaborate upon judgemental sampling and snowball sampling.

7.11 FURTHER READINGS

- Saravanel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.
- Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.
- Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.
- Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.
- Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

*Sampling Theory:
Design, Size and
Techniques*

NOTES

UNIT 8 HYPOTHESIS TESTING

NOTES

Structure

- 8.0 Introduction
- 8.1 Objectives
- 8.2 Hypothesis
- 8.3 Steps Involved in Hypothesis Testing
- 8.4 Parametric and Non-Parametric Tests
- 8.5 T-Distribution
- 8.6 Tests of Significance of a Mean, Standard Deviation and Proportion
 - 8.6.1 Test of Significance of Means - Case of Single Population
 - 8.6.2 Tests of Significance of Difference between two Population Means
 - 8.6.3 Test of Significance of Sample Population Proportion
 - 8.6.4 Tests Significance of Difference between two Population Proportions
- 8.7 Answers to Check Your Progress Questions
- 8.8 Summary
- 8.9 Key Words
- 8.10 Self Assessment Questions and Exercises
- 8.11 Further Readings

8.0 INTRODUCTION

Hypothesis testing is a process whereby a researcher or analyst tests the plausibility of an assumption regarding a population parameter. The process follows a series of steps for the purpose. It begins with establishing a hypothesis and is followed by the establishment of a suitable significance level, determination of a test statistic, determination of critical region, computation of the value of a test-statistic and making a decision. A hypothesis test is important as it evaluates different statements to determine which one is best supported by the sample data. This unit focuses on the nuances of hypothesis testing and the methods used for the same.

8.1 OBJECTIVES

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

- Understand the steps involved in Hypothesis Testing
- Discuss Parametric and Non-Parametric tests
- Analyze T-distribution
- Describe the tests of significance of Mean, Standard Deviation and Proportion

8.2 HYPOTHESIS

For an investigation to be fruitful, one should collect such data or information for or against some point of ‘view’ or ‘proposition’. Such a view point or proposition is termed as *hypothesis*. Hypothesis is a proposition temporarily accepted as true in the light of what is, at the time, known about a phenomenon. In the words of *Lunberg* a hypothesis is “a tentative generalization the validity of which to be tested”. Thus, hypothesis is any guess, imaginative idea which becomes a basis for action or investigation.

When to Make a Hypothesis

Hypotheses are not given to us readymade. This is true especially in social sciences where there has not yet evolved a highly developed theoretical system in many areas of its subject matter, affording fruitful bases for hypothesis formulation. As a result, in social sciences at least, a considerable portion of research endeavour is directed understandably towards making hypotheses rather than testing them. Hence, it is to be remembered that research can begin with well formulated hypothesis (in the planning stage of the project) or it can come out with hypothesis as its end product (at a later stage).

Need for Hypothesis

While a hypothesis is useful it is not always indispensable. In physical sciences (including fundamental research), the hypothesis is most often necessary but in social sciences, useful facts may be discovered, organized and presented purposefully even without a hypothesis other than the one which might incidentally be suggested in the course of the investigation. Thus it is desirable to have hypothesis / hypotheses.

Types of Hypothesis

There can be two basic type of hypothesis, viz., (i) descriptive and (ii) relational.

- (i) **Descriptive Hypothesis:** Descriptive hypothesis concerns the existence, size, form or distribution of some concept subject to verification. Concepts presented may take on two or more values. For example, “over half of Golden Eagle Beer sales is to the university students”. Descriptive hypothesis in question form becomes, “Is over half of Golden Eagle Beer sales to university students?”
- (ii) **Relational Hypothesis:** Relational hypothesis is proposed in order to test the relationship or linkage between two variables. For example, “High consumption of Golden Eagle Beer among university students is a result of heavy sales promotion directed to this market segment.” This involves a relationship between two variables, namely, consumption and sales promotion.

NOTES

NOTES

The nature of the relationship between the two variables is not specified. Is there only an implication that the variables occur in some predictable relationship, or is one variable somehow responsible for the other? The first interpretation (unspecified relationship) indicates a correlational relationship; the second (predictable relationship) indicates an explanatory, or causal, relationship.

Correlational relationships state merely that the variables occur together in some specified manner without implying that one causes the other. Such weak claims are often made when we believe there are more basic causal forces that affect both variables or when we have not developed enough evidence to claim a stronger linkage. Here are three sample correlational hypotheses : (1) Young machinists (under 35 years of age) are less productive than those who are 35 years or older. (2) The height of women's hemlines varies directly with the level of the business cycle. (3) People in give the president a more favorable rating than do people in Pakistan. By labeling these as correlational hypotheses, we make no claim that one variable causes the other to change or take on different values. Other persons, however, may view one or more of these hypotheses as reflecting cause-and effect relationships.

With **explanatory, (causal), hypotheses**, there is an implication that the existence of, or a change in, one variable causes or leads to an effect on the other variable. As we noted previously, the causal variable is typically called the *independent variable* (IV) and the other the *dependent variable* (DV). Cause means roughly to "help make happen." So, the independent variable need not be the sole reason for the existence of, or change in, the dependent variable. Here are few examples of explanatory hypotheses:

- An increase in family income leads to an increase in the percentage of income saved.
- Exposure to the company's messages concerning industry problems leads to more favorable attitudes by production workers toward the company.
- Loyalty to a particular grocery store increases the probability of purchasing the private brands sponsored by that store.
- An increase in the price of salvaged copper wire leads to an increase in scavenging on the Army firing range.

Sources of Hypothesis

It goes without saying that the formulation and verification of hypothesis is objective of any scientific investigation. A hypothesis may be developed from the following.

1. The scientific theory which is available for putting for validity. For example, a flourishing market has emerged for readymades. Another example is a Good market for good products.
2. The findings of previous marketing surveys lead towards the formulation of hypothesis. A researcher may hypothesize the earlier works that have similar or dissimilar marketing trends.

3. The culture or opinion of consumers has a great impact upon the thinking process. In finding solutions to value-oriented problems, one cannot lose sight of culture. The cultural emphasis on happiness due to good marketing conditions has been productive.
4. Likeness and similarity result in operation of the view point. For example, if two products are similar in quality as well as in measurement, but differ in price. Here the question arises why.
5. The personal experiences and individual reactions in marketing are good source of hypothesis. It is an experienced person who sees the invisible phenomena of marketing conditions in the right perspective and formulates a hypothesis.
6. A second source of hypotheses is folk wisdom or current popular beliefs and practices suggesting both the problems and the hypothesis.
7. Relevant and available techniques: An investigation, in order to be practical, should relate the hypothesis to the investigational techniques feasible in that particular science. This requires in the researcher a sound knowledge of techniques to test the hypothesis and, thus, to formulate practical questions.

NOTES

Check Your Progress

1. What is descriptive hypothesis concerned with?
2. Why is relational hypothesis proposed?

8.3 STEPS INVOLVED IN HYPOTHESIS TESTING

The following fundamental steps are followed in testing of a hypothesis:

Setting up of a hypothesis: First step is to establish the hypothesis to be tested. As it is known, these statistical hypotheses are generally assumptions about the value of the population parameter; the hypothesis specifies a single value or a range of values for two different hypotheses rather than constructing a single hypothesis. These two hypotheses are generally referred to as the (1) null hypotheses denoted by H_0 and (2) alternative hypothesis denoted by H_1 .

The null hypothesis is the hypothesis of the population parameter taking a specified value. In case of two populations, the null hypothesis is of no difference or the difference taking a specified value. The hypothesis that is different from the null hypothesis is the alternative hypothesis. If the null hypothesis H_0 is rejected based upon the sample information, the alternative hypothesis H_1 is accepted. Therefore, the two hypotheses are constructed in such a way that if one is true, the other one is false and vice versa. There can also be situations where the researcher is interested in establishing the relationship between any two variables. In such a case, a null hypothesis is set as the hypothesis of no relationship between those two variables; whereas the alternative hypothesis is the hypothesis of the relationship

NOTES

between variables. The rejection of the null hypothesis indicates that the differences/relationship have a statistical significance and the acceptance of the null hypothesis means that any difference/relationship is due to chance.

Setting up of a suitable significance level: The next step in the testing of hypothesis exercise is to choose a suitable level of significance. The level of significance denoted by $\pm$ is chosen before drawing any sample. The level of significance denotes the probability of rejecting the null hypothesis when it is true. The value of $\pm$ varies from problem to problem, but usually it is taken as either 5 per cent or 1 per cent. A 5 per cent level of significance means that there are 5 chances out of hundred that a null hypothesis will get rejected when it should be accepted. This means that the researcher is 95 per cent confident that a right decision has been taken. Therefore, it is seen that the confidence with which a researcher rejects or accepts a null hypothesis depends upon the level of significance. When the null hypothesis is rejected at any level of significance, the test result is said to be significant. Further, if a hypothesis is rejected at 1 per cent level, it must also be rejected at 5 per cent significance level.

Determination of a test statistic: The next step is to determine a suitable test statistic and its distribution. As would be seen later, the test statistic could be t , Z , χ^2 or F , depending upon various assumptions to be discussed later in the book.

Determination of critical region: Before a sample is drawn from the population, it is very important to specify the values of test statistic that will lead to rejection or acceptance of the null hypothesis. The one that leads to the rejection of null hypothesis is called the critical region. Given a level of significance, $\pm$, the optimal critical region for a two-tailed test consists of that $\pm/2$ per cent area in the right hand tail of the distribution plus that $\pm/2$ per cent in the left hand tail of the distribution where that null hypothesis is rejected. Therefore, establishing a critical region is similar to determining a 100 ($1 - \pm$) per cent confidence interval.

Computing the value of test-statistic: The next step is to compute the value of the test statistic based upon a random sample of size n . Once the value of test statistic is computed, one needs to examine whether the sample results fall in the critical region or in the acceptance region.

Making decision: The hypothesis may be rejected or accepted depending upon whether the value of the test statistic falls in the rejection or the acceptance region. Management decisions are based upon the statistical decision of either rejecting or accepting the null hypothesis.

If the hypothesis is being tested at 5 per cent level of significance, it would be rejected if the observed results have a probability less than 5 per cent. In such a case, the difference between the sample statistic and the hypothesized population parameter is considered to be significant. On the other hand, if the hypothesis is

accepted, the difference between the sample statistic and the hypothesized population parameter is not regarded as significant and can be attributed to chance.

Check Your Progress

3. What are the two types of hypothesis referred to as?
4. What do the rejection and acceptance of a null hypothesis indicate?
5. Why may a hypothesis be rejected?

NOTES

8.4 PARAMETRIC AND NON-PARAMETRIC TESTS

A hypothesis is tested to identify the errors that have occurred in the statements and concepts used in that hypothesis. Hypothesis testing can be broadly divided into two types, which are as follows:

- Parametric tests or standard tests of hypothesis
- Non-parametric tests or distribution-free tests of hypothesis

The test on the parameters like mean, standard deviation and proportion are called parametric tests. Parametric tests are more powerful because their data is either of interval or ratio level and based on the following assumptions:

- (a) The observations must be independent.
- (b) The observations need to be drawn from populations that are normally distributed.
- (c) The populations should have equal variances.

It is the researcher's responsibility to check the assumptions relevant to the chosen test. Some of the popular parametric tests are z-test, t-test and F-test.

Non-parametric tests, on the other hand, have less stringent and fewer assumptions. Non-parametric tests are called distribution free tests as they do not require any assumption regarding the shape of the population distribution from where the sample is drawn. They do not specify normally distributed populations or equality of variances. Some tests require independence of cases, while others are designed expressly for situations with related cases. Non-parametric tests are generally used for qualitative analysis (ordinal or nominal-level data). Both the categories of tests provide efficient results provided their selections are appropriate. Non-parametric tests include chi-square, run-test, Mann-Whitney test, Kruskal-Wallis test, etc.

1. Parametric tests or standard tests of hypothesis

Parametric tests assume certain properties of the population sample such as observations from a normal population, large sample size, population parameters

NOTES

like mean and variance. The various parametric tests of hypothesis are based on the assumption of normality. In other words, the source of data for them is normally distributed. They can be listed as follows:

- **Z-test:** This kind of test is based on normal probability distribution. It is mostly used to judge the significance of mean as a statistical measure. This is the most frequently used test in research studies. It is generally used to compare the mean of a sample with the hypothesized mean of the population. It is also used in case the population variance is known. It is helpful in judging the significance of difference between the means of two independent large samples, to compare the sample proportion to a theoretical value of population proportion and to judge the significance of median, mode and coefficient of correlation.
- **T-test:** This test is based on t-distribution and is used to judge the significance of a sample mean or the difference between the means of two small samples when the population variance is not known.
- χ^2 : This test is based on a chi-square distribution and is used for comparing a sample variance to a theoretical population variance.
- **F-test:** This test is based on F-distribution and is also used to compare the variance of two independent samples. It is also used to compare the significance of multiple correlation coefficients.

2. *Non-parametric tests or distribution-free tests of hypothesis*

There are situations in testing where assumptions cannot be made. In such situations, non-parametric methods are employed. There are various types of non-parametric tests. The important ones are as follows:

- **Sign test:** This is one of the easiest tests in practice based on the plus/minus sign of an observation in a sample. The sign may be one of the following two types:
 - o **One-sample sign test:** This is a very simple distribution-free test and is applied in case of a sample from a continuous symmetrical population, wherein the probability of a sample to be either less or more than mean is half. Here, to test a null hypothesis, all those items, which are greater than the alternate hypothesis, are replaced by a plus sign and those, which are less than the alternate hypothesis, are replaced by a minus sign.
 - o **Two-sample sign test:** In case of all the problems consisting of paired data, two-sample sign test is used. Here, each pair of values can be replaced with a plus sign in the first value of the first sample with the first value of the second sample. If the first value is less, minus sign is assigned.
- **Fisher–Irwin test:** This is applied where there is no difference between two sets of data. In other words, it is used where you can assume that two

different treatments are supposedly different in terms of the results that they produce. It is applied in all those cases where result for each item in a sample can be divided into one of the two mutually exclusive categories.

- **McNamara test:** It is applied where the data is nominal in nature and is related to two interrelated samples. By using this test, you can judge the significance of any observed changes in the same subject.
- **Wilcoxon matched-pairs test:** This test is applied in the case of a matched-pair such as output of two similar machines. Here, you can determine both the direction and the magnitude between the matched values. This test is also called Signed Rank Test.

NOTES

Check Your Progress

6. List the two types of hypothesis testing.
7. Name some of the popular parametric tests.
8. What is Z-test helpful in doing?

8.5 T-DISTRIBUTION

The method for comparing two related samples in hypothesis testing is the paired **t-test**. For this test, it is necessary that the observations in the two samples should be collected in the form of matched pairs. It means that each observation in one sample must be paired with an observation in the other sample in such a way that they are matched excluding the other factors which do not fall within the area of this test.

An important formula for calculating this is:

$$\bar{D} = \frac{\sum D_i}{n}$$

And the formula for calculating the variances of the differences is:

$$(\sigma_{\text{diff}})^2 = \frac{\sum D_i^2 - (\bar{D})^2 \times n}{n - 1}$$

Where, $\bar{D}$ = Mean of differences.

The t-test is based on t-distribution, which is a probability distribution that arises in the problem of estimating the mean of a normally distributed population when the sample size is small. t-distribution arises when the population standard deviation is unknown and has to be estimated from the data.

In Student's t-distribution, the random variable is assumed as X with mean μ and variance σ^2 , $Z_{\bar{x}}$, the standard normal statistic is assumed as $\bar{X}$ and X^2

NOTES

be a random variable, which follows chi-square distribution with t degrees of freedom.

If the variables are relatively independent with each other, then the t-distribution will be:

$$t = \frac{Z_x}{\sqrt{\frac{X^2}{t}}}$$

The standard normal statistic and chi-square statistics are:

$$Z_x = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

and

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \text{ with } (n-1) \text{ degrees of freedom.}$$

After substituting these two statistics in t, t-distribution is represented as:

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}} \text{ with } (n-1) \text{ degrees of freedom.}$$

The t-distribution can be used only when the sample size is not more than 30; when the sample size is more than 30, then it can be approximated to a normal distribution.

8.6 TESTS OF SIGNIFICANCE OF A MEAN, STANDARD DEVIATION AND PROPORTION

In this section, you will learn about tests of significance of mean, standard deviation and proportion.

Tests Concerning Population Mean

In this section, we will take up the test of hypothesis about population mean in a case of single population and the difference between the two means for two populations. One of the important things that have to be kept in mind is the use of an appropriate test statistic. In case the sample size is large ($n > 30$), Z statistic would be used. For a small sample size ($n \leq 30$), a further question regarding the knowledge of population standard deviation (σ) is asked. If the population standard deviation σ is known, a Z statistic can be used. However, if σ is unknown and is estimated using sample data, a t test with appropriate degrees of freedom is used under the assumption that the sample is drawn from a normal population. It is

assumed that the readers have the knowledge of Z and t distribution from the course on statistics. However, these would be briefly reviewed at the appropriate place.

Tests Concerning Population Proportion

In the tests about proportion, one is interested in examining whether the respondents possess a particular attribute or not.

The random variable in such a case is a binary one in the sense it takes only two values—yes or no. As we know that either a student is a smoker or not, a consumer either uses a particular brand of product or not and lastly, a skilled worker may be either satisfied or not with the present job. At this stage it may be recalled that the binomial distribution is a theoretically correct distribution to use while dealing with proportions. Further, as the sample size increases, the binomial distribution approaches the normal distribution in characteristic. To be specific, whenever both np and nq (where n = number of trials, p = probability of success and q = probability of failure) are at least 5, one can use the normal distribution as a substitute for the binomial distribution.

8.6.1 Test of Significance of Means - Case of Single Population

In this section, a number of illustrations will be taken up to explain the test of hypothesis concerning mean. Two cases of large sample and small samples will be taken up.

(i) Case of Large Sample

As mentioned earlier, in case the sample size n is large or small but the value of the population standard deviation is known, a Z test is appropriate. There can be alternate cases of two-tailed and one-tailed tests of hypotheses.

Corresponding to the null hypothesis $H_0 : \mu = \mu_0$, the following criteria could be used as shown in Table 8.1.

The test statistic is given by,

$$Z = \frac{\bar{X} - \mu_{H_0}}{\frac{\sigma}{\sqrt{n}}}$$

Where,

$\bar{X}$ = Sample mean

σ = Population standard deviation

μ_{H_0} = The value of μ under the assumption that the null hypothesis is true.

n = Size of sample.

NOTES

Table 8.1 Criteria for Accepting or Rejecting Null Hypothesis under Different Cases of Alternative Hypotheses**NOTES**

S. No.	Alternative Hypothesis	Reject the Null Hypothesis if	Accept the Null Hypothesis if
1.	$\mu < \mu_0$	$Z < -Z_\alpha$	$Z \geq -Z_\alpha$
2.	$\mu > \mu_0$	$Z > Z_\alpha$	$Z \leq Z_\alpha$
3.	$\mu \neq \mu_0$	$Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$	$-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}$

If the population standard deviation σ is unknown, the sample standard

$$\text{deviation } s = \sqrt{\frac{1}{n-1} \sum (X - \bar{X})^2}$$

is used as an estimate of σ . It may be noted that Z_α and $Z_{\alpha/2}$ are Z values such that the area to the right under the standard normal distribution is α and $\alpha/2$ respectively. Below are solved examples using the above concepts.

Example 8.1: A sample of 200 bulbs made by a company give a lifetime mean of 1540 hours with a standard deviation of 42 hours. Is it likely that the sample has been drawn from a population with a mean lifetime of 1500 hours? You may use 5 per cent level of significance.

Solution:

In the above example, the sample size is large ($n = 200$), sample mean ($\bar{X}$) equals 1540 hours and the sample standard deviation (s) is equal to 42 hours. The null and alternative hypotheses can be written as:

$$H_0 : \mu = 1500 \text{ hrs}$$

$$H_1 : \mu \neq 1500 \text{ hrs}$$

It is a two-tailed test with level of significance (α) to be equal to 0.05. Since n is large ($n > 30$), though population standard deviation σ is unknown, one can use Z test. The test statistics are given by:

$$Z = \frac{\bar{X} - \mu_{H0}}{\frac{\sigma}{\sqrt{n}}}$$

Where, μ_{H0} = Value of μ under the assumption that the null hypothesis is true

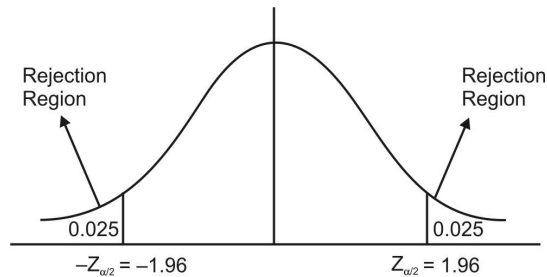
$\frac{\sigma}{\sqrt{n}}$ = Estimated standard error of mean

$$\text{Here } \mu_{H0} = 1,500, \frac{\sigma}{\sqrt{n}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{42}{\sqrt{200}} = 2.97$$

(Note that $\hat{\sigma}$ is estimated value of σ .)

$$Z = \frac{\bar{X} - \mu_{H0}}{\frac{s}{\sqrt{n}}} = \frac{1,540 - 1,500}{2.97} = \frac{40}{2.97} = 13.47$$

The value of $\alpha = 0.05$ and since it is a two-tailed test, the critical value Z is given by $-Z_{\alpha/2}$ and $Z_{\alpha/2}$ which could be obtained from the standard normal Table 7.1 given in Unit 7.



Since the computed value of $Z = 13.47$ lies in the rejection region, the null hypothesis is rejected. Therefore, it can be concluded that the average life of the bulb is significantly different from 1,500 hours.

Example 8.2: On a typing test, a random sample of 36 graduates of a secretarial school averaged 73.6 words with a standard deviation of 8.10 words per minute. Test an employer's claim that the school's graduates average less than 75.0 words per minute using the 5 per cent level of significance.

Solution:

$$H_0 : \mu = 75$$

$$H_1 : \mu < 75$$

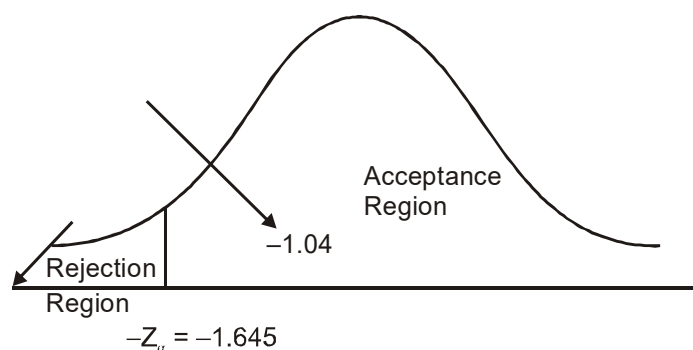
$\bar{X} = 73.6$, $s = 8.10$, $n = 36$ and $\alpha = 0.05$. As the sample size is large ($n > 30$), though population standard deviation σ is unknown, Z test is appropriate.

The test statistic is given by:

$$Z = \frac{\bar{X} - \mu_{H0}}{\frac{\hat{\sigma}}{\sqrt{n}}} = \frac{73.6 - 75}{\frac{8.10}{\sqrt{36}}} = \frac{-1.4}{1.35} = -1.04$$

$$\left(\frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{8.10}{\sqrt{36}} = \frac{8.10}{6} = 1.35 \right)$$

Since it is a one-tailed test and the interest is in the left hand tail of the distribution, the critical value of Z is given by $-Z_{\alpha} = -1.645$. Now, the computed value of Z lies in the acceptance region, and the null hypothesis is accepted as shown below:



NOTES

NOTES

(ii) Case of Small Sample

In case the sample size is small ($n \leq 30$) and is drawn from a population having a normal population with unknown standard deviation σ , a t test is used to conduct the hypothesis for the test of mean. The t distribution is a symmetrical distribution just like the normal one. However, t distribution is higher at the tail and lower at the peak. The t distribution is flatter than the normal distribution. With an increase in the sample size (and hence degrees of freedom), t distribution loses its flatness and approaches the normal distribution whenever $n > 30$. A comparative shape of t and normal distribution is given in Figure 8.1.

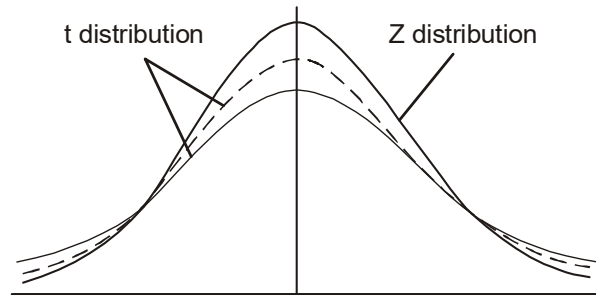


Fig. 8.1 Shape of t and Normal Distribution

The procedure for testing the hypothesis of a mean is similar to what is explained in the case of large sample. The test statistic used in this case is:

$$t_{n-1} = \frac{\bar{X} - \mu_{H0}}{\frac{\hat{\sigma}}{\sqrt{n}}}$$

Where, $\frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n}}$ (where s = Sample standard deviation)

$n - 1$ = degrees of freedom

A few examples pertaining to ' t ' test are worked out for testing the hypothesis of mean in case of a small sample.

Example 8.3: Prices of share (in ₹) of a company on the different days in a month were found to be 66, 65, 69, 70, 69, 71, 70, 63, 64 and 68. Examine whether the mean price of shares in the month is different from 65. You may use 10 per cent level of significance.

Solution:

$$H_0 : \mu = 65$$

$$H_1 : \mu \neq 65$$

Since the sample size is $n = 10$, which is small, and the sample standard deviation is unknown, the appropriate test in this case would be t . First of all, we need to estimate the value of sample mean ($\bar{X}$) and the sample standard deviation

(s). It is known that the sample mean and the standard deviation are given by the following formula.

$$\bar{X} = \frac{\sum X}{n} \quad s = \sqrt{\frac{1}{n-1} \sum (X - \bar{X})^2}$$

The computation of $\bar{X}$ and s is shown in Table 10.3.

$$\sum X = 675, \quad \bar{X} = \frac{\sum X}{n} = \frac{675}{10} = 67.5$$

$$\sum (X - \bar{X})^2 = 70.5$$

$$s^2 = \frac{1}{n-1} \sum (X - \bar{X})^2 = \frac{70.5}{9} = 7.83$$

$$s = \sqrt{7.83} = 2.80$$

Table 8.2 Computation of Sample Mean and Standard Deviation

S. No.	X	$X - \bar{X}$	$(X - \bar{X})^2$
1	66	-1.5	2.25
2	65	-2.5	6.25
3	69	1.5	2.25
4	70	2.5	6.25
5	69	1.5	2.25
6	71	3.5	12.25
7	70	2.5	6.25
8	63	-4.5	20.25
9	64	-3.5	12.25
10	68	0.5	0.25
Total	675	0	70.5

The test statistic is given by:

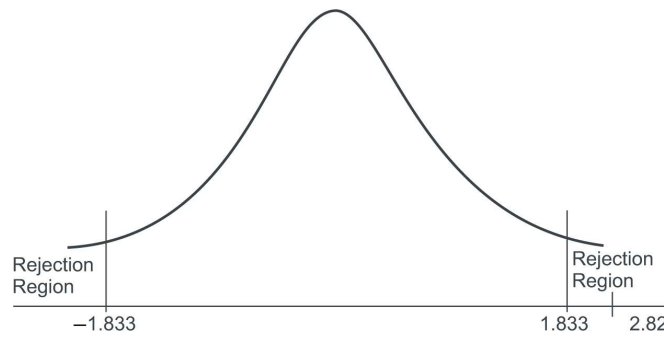
$$t_{n-1} = \frac{\frac{\bar{X} - \mu_{H0}}{\frac{\hat{\sigma}}{x}}}{\frac{s}{\sqrt{n}}} = \frac{\bar{X} - \mu_{H0}}{\frac{s}{\sqrt{n}}} = \frac{67.5 - 65}{\frac{2.8}{\sqrt{10}}} = \frac{2.5 \times \sqrt{10}}{2.8}$$

$$= 2.5 \times 3.16 / 2.8 = 7.91 / 2.8 = 2.82$$

The critical values of t with 9 degrees of freedom for a two-tailed test are given by -1.833 and 1.833. Since the computed value of t lies in the rejection region (see figure below), the null hypotheses is rejected.

NOTES

NOTES



Therefore, the average price of the share of the company is different from 65.

Example 8.4: Past records indicate that a golfer has averaged 82 on a certain course. With a new set of clubs, he averages 79 over five rounds with a standard deviation of 2.65. Can we conclude that at 0.025 level of significance, the new club has an adverse effect on the performance?

Solution:

$$H_0 : \mu = 82$$

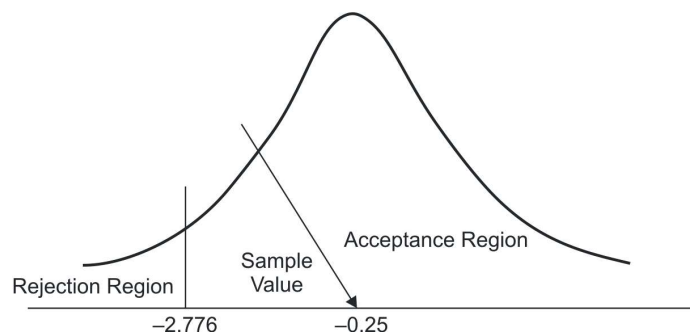
$$H_1 : \mu < 82$$

$\bar{X} = 79, n = 5, s = 2.65, \alpha = 0.025$. As the population standard deviation is unknown and the sample size is small ($n < 30$), a t test would be appropriate. The test statistic is given by:

$$t_{n-1} = \frac{\bar{X} - \mu_{H_0}}{\frac{s}{\sqrt{n}}} = \frac{\bar{X} - \mu_{H_0}}{s/\sqrt{n}} = \frac{79 - 82}{1.185} = \frac{-0.3}{1.185} = -0.25$$

$$\left(\frac{s}{\sqrt{n}} = \frac{2.65}{\sqrt{5}} = 1.185 \right)$$

The critical value of t at 0.025 level of significance with four degrees of freedom is given by $-t_{\alpha} = -2.776$ (see Table 8.3). As the sample t value of -0.25 lies in the acceptance region, the null hypothesis is accepted (see figure below).



Therefore, there is no adverse effect on the performance due to a change in the club and the performance can be attributed to chance.

Table 8.3 Some critical values of 't'

Degrees of Freedom	Level of Significance		
	1%	5%	10%
1	63.657	12.706	6.314
2	9.925	4.303	2.920
3	5.841	3.182	2.353
4	4.604	2.776	2.132
5	4.032	2.571	2.015
6	3.707	2.447	1.943
7	3.499	2.365	1.895
8	3.355	2.306	1.860
9	3.250	2.262	1.833
10	3.169	2.228	1.812
11	3.106	2.201	1.796
12	3.055	2.179	1.782
13	3.012	2.160	1.771
14	2.977	2.145	1.761
15	2.947	2.131	1.753
16	2.921	2.120	1.746
17	2.898	2.110	1.740
18	2.878	2.101	1.734
19	2.861	2.093	1.729
20	2.845	2.086	1.725
21	2.831	2.080	1.721
22	2.819	2.074	1.717
23	2.807	2.069	1.714
24	2.797	2.064	1.711
25	2.787	2.060	1.708
26	2.779	2.056	1.706
27	2.771	2.052	1.703
28	2.763	2.048	1.701
29	2.756	2.045	1.699
α	2.576	1.960	1.645

NOTES

Note: These table values of 't' are in respect of two-tailed tests. If we use the *t*-distribution for one-tailed test then we are interested in determining the area located in one tail. So to find the appropriate *t*-value for a one-tailed test say at a 5% level with 12 degrees of freedom, then we should look in the above table under the 10% column opposite the 12 degrees of freedom row. (This value will be 1.782). This is true because the 10% column represents 10% of the area under the curve contained in both tails combined, and so it also represents 5% of the area under the curve contained in each of the tails separately.

8.6.2 Tests of Significance of Difference between two Population Means

So far, we have been concerned with the testing of means of a single population. We took up the cases of both large and small samples. It would be interesting to examine the difference between the two population means. Again, various cases would be examined as discussed follows:

NOTES**(i) Case of Large Sample**

In case both the sample sizes are greater than 30, a Z test is used. The hypothesis to be tested may be written as:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Where,

μ_1 = mean of population 1

μ_2 = mean of population 2

The above is a case of two-tailed test. The test statistic used is:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$\bar{X}_1$ = Mean of sample drawn from population 1

$\bar{X}_2$ = Mean of sample drawn from population 2

n_1 = size of sample drawn from population 1

n_2 = size of sample drawn from population 2

If σ_1 and σ_2 are unknown, their estimates given by s_1 and s_2 are used.

$$\hat{\sigma}_1 = s_1 = \sqrt{\frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2}$$

$$\hat{\sigma}_2 = s_2 = \sqrt{\frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2}$$

The Z value for the problem can be computed using the above formula and compared with the table value to either accept or reject the hypothesis. Let us consider the following problem:

Example 8.5: A study is carried out to examine whether the mean hourly wages of the unskilled workers in the two cities—Ambala Cantt and Lucknow are the same. The random sample of hourly earnings in both the cities is taken and the results are presented in the Table 8.4.

Table 8.4 Survey Data on Hourly Earnings in Two Cities

City	Sample Mean Hourly Earnings	Standard Deviation of Sample	Sample Size
Ambala Cantt	₹ 8.95 ($\bar{X}_1$)	0.40 (s_1)	200 (n_1)
Lucknow	₹ 9.10 ($\bar{X}_2$)	0.60 (s_2)	175 (n_2)

Using a 5 per cent level of significance, test the hypothesis of no difference in the average wages of unskilled workers in the two cities.

Solution: We use subscripts 1 and 2 for Ambala Cantt and Lucknow respectively.

$$H_0 : \mu_1 = \mu_2 \rightarrow \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2 \rightarrow \mu_1 - \mu_2 \neq 0$$

The following survey data is given:

$$\bar{X}_1 = 8.95, \bar{X}_2 = 9.10, s_1 = 0.40, s_2 = 0.60, n_1 = 200, n_2 = 175, \alpha = 0.05$$

Since both n_1, n_2 are greater than 30 and the sample standard deviations are given, a Z test would be appropriate.

The test statistic is given by:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)H_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

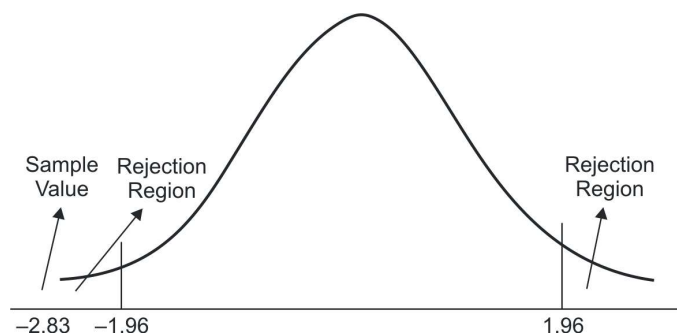
As σ_1, σ_2 are unknown, their estimates would be used.

$$s_1 = \hat{\sigma}_1, \quad s_2 = \hat{\sigma}_2$$

$$\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} = \sqrt{\frac{(0.4)^2}{200} + \frac{(0.6)^2}{175}} = \sqrt{0.0028} = 0.0053$$

$$Z = \frac{(8.95 - 9.10) - 0}{0.0053} = -2.83$$

As the problem is of a two-tailed test, the critical values of Z at 5 per cent level of significance are given by $-Z_{\alpha/2} = -1.96$ and $Z_{\alpha/2} = 1.96$. The sample value of $Z = -2.83$ lies in the rejection region as shown in the figure below:



(ii) Case of Small Sample

If the size of both the samples is less than 30 and the population standard deviation is unknown, the procedure described above to discuss the equality of two

NOTES

NOTES

population means is not applicable in the sense that a t test would be applicable under the assumptions:

- (a) Two population variances are equal.
- (b) Two population variances are not equal.

(a) Population variances are equal

If the two population variances are equal, it implies that their respective unbiased estimates are also equal. In such a case, the expression becomes:

$$\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} = \sqrt{\frac{\hat{\sigma}^2}{n_1} + \frac{\hat{\sigma}^2}{n_2}} = \hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

(Assuming $\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = \hat{\sigma}^2$)

To get an estimate of $\hat{\sigma}^2$, a weighted average of s_1^2 and s_2^2 is used, where the weights are the number of degrees of freedom of each sample. The weighted average is called a 'pooled estimate' of $\hat{\sigma}^2$. This pooled estimate is given by the expression:

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

The testing procedure could be explained as under:

$$H_0 : \mu_1 = \mu_2 \rightarrow \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2 \rightarrow \mu_1 - \mu_2 \neq 0$$

In this case, the test statistic t is given by the expression:

$$t_{n_1+n_2-2} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)H_0}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where,

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Once the value of t statistic is computed from the sample data, it is compared with the tabulated value at a level of significance α to arrive at a decision regarding the acceptance or rejection of hypothesis. Let us work out a problem illustrating the concepts defined above.

Example 8.6: Two drugs meant to provide relief to arthritis sufferers were produced in two different laboratories. The first drug was administered to a group of 12 patients and produced an average of 8.5 hours of relief with a standard deviation of 1.8 hours. The second drug was tested on a sample of 8 patients and produced an average of 7.9 hours of relief with a standard deviation of 2.1 hours.

Test the hypothesis that the first drug provides a significantly higher period of relief. You may use 5 per cent level of significance.

Solution: Let the subscripts 1 and 2 refer to drug 1 and drug 2 respectively.

$$H_0 : \mu_1 = \mu_2 \rightarrow \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2 \rightarrow \mu_1 - \mu_2 \neq 0$$

The following survey data is given:

$$\bar{X}_1 = 8.5, \bar{X}_2 = 7.9, s_1 = 1.8, s_2 = 2.1, n_1 = 12, n_2 = 8$$

As both n_1, n_2 are small and the sample standard deviations are unknown, one may use a t test with the degrees of freedom $= n_1 + n_2 - 2 = 12 + 8 - 2 = 18$ d.f.

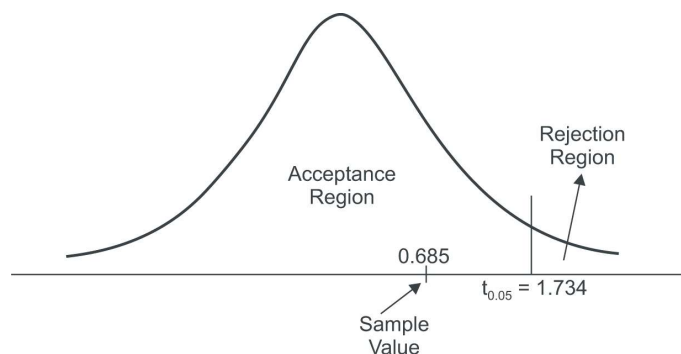
The test statistics is given by:

$$t_{n_1+n_2-2} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)H_0}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where,

$$\begin{aligned} \hat{\sigma} &= \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \\ &= \sqrt{\frac{(12 - 1)(1.8)^2 + (8 - 1)(2.1)^2}{12 + 8 - 2}} = \sqrt{\frac{11 \times 3.24 + 7 \times (4.41)}{18}} \\ &= \sqrt{\frac{35.64 + 30.87}{18}} = \sqrt{\frac{66.61}{18}} = \sqrt{3.698} = 1.92 \\ t_{18} &= \frac{(8.5 - 7.9) - (0)}{1.92 \sqrt{\frac{1}{12} + \frac{1}{8}}} = \frac{0.6}{1.92 \sqrt{0.2083}} \\ &= \frac{0.6}{1.92 \times 0.456} = \frac{0.6}{0.8755} = 0.685 \end{aligned}$$

The critical value of t with 18 degrees of freedom at 5 per cent level of significance is given by 1.734. The sample value of $t = 0.685$ lies in the acceptance region as shown in figure below:



NOTES

Therefore, the null hypothesis is accepted as there is not enough evidence to reject it. Therefore, one may conclude that the first drug is not significantly more effective than the second drug.

NOTES

(b) When population variances are not equal

In case population variances are not equal, the test statistic for testing the equality of two population means when the size of samples are small is given by:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)H_0}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}}$$

The degrees of freedom in such a case is given by the expression:

$$d.f = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2}$$

The procedure for testing of hypothesis remains the same as was discussed when the variances of two populations were assumed to be same. Let us consider an example to illustrate the same.

Example 8.7: There were two types of drugs (1 and 2) that were tried on some patients for reducing weight. There were 8 adults who were subjected to drug 1 and seven adults who were administered drug 2. The decrease in weight (in pounds) is given below:

Drug 1	10	8	12	14	7	15	13	11
Drug 2	12	10	7	6	12	11	12	

Do the drugs differ significantly in their effect on decreasing weight? You may use 5 per cent level of significance. Assume that the variances of two populations are not same.

Solution:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Let us compute the sample means and standard deviations of the two samples as shown in Table 8.5.

Table 8.5 Intermediate computations for sample means and standard deviations

Hypothesis Testing

S.No.	X_1	X_2	$(X_1 - \bar{X}_1)$	$(X_2 - \bar{X}_2)$	$(X_1 - \bar{X}_1)^2$	$(X_2 - \bar{X}_2)^2$
1	10	12	-1.25	2	1.5625	4
2	8	10	-3.25	0	10.5625	0
3	12	7	0.75	-3	0.5625	9
4	14	6	2.75	-4	7.5625	16
5	7	12	-4.25	2	18.0625	4
6	15	11	3.75	1	14.0625	1
7	13	12	1.75	2	3.0625	4
8	11		-0.25		0.0625	
Total	90	70	0	0	55.5	38
Mean	11.25	10				

NOTES

$$n_1 = 8,$$

$$n_2 = 7,$$

$$\bar{X}_1 = \frac{\sum X_1}{n_1} = \frac{90}{8} = 11.25$$

$$\bar{X}_2 = \frac{\sum X_2}{n_2} = \frac{70}{7} = 10$$

$$s_1^2 = \frac{\sum (X_1 - \bar{X}_1)^2}{n_1 - 1} = \frac{55.5}{7} = 7.93$$

$$s_2^2 = \frac{\sum (X_2 - \bar{X}_2)^2}{n_2 - 1} = \frac{38}{6} = 6.33$$

$$\hat{\sigma}_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{7.93}{8} + \frac{6.33}{7}} = \sqrt{0.99 + 0.90} = \sqrt{1.89} = 1.37$$

$$d.f. = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right) + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)} = \frac{\left(\frac{7.33}{8} + \frac{6.33}{7}\right)^2}{\frac{1}{7} \left(\frac{7.33}{8}\right)^2 + \frac{1}{6} \left(\frac{6.33}{7}\right)^2}$$

$$= \frac{3.314}{0.12 + 0.136} = \frac{3.314}{0.12 + 0.136} = 12.996 = 13 \text{ (approx.)}$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)H_0}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}}$$

$$t = \frac{11.25 - 10}{1.37} = \frac{1.25}{1.37} = 0.912$$

The table value (critical value) of t with 13 degrees of freedom at 5 per cent level of significance is given by 2.16. As computed t is less than tabulated t , there is not enough evidence to reject H_0 .

8.6.3 Test of Significance of Sample Population Proportion

Suppose we want to test the hypotheses,

$$H_0 : p = p_0$$

$$H_1 : p \neq p_0$$

For large sample, the appropriate test statistic would be:

$$Z = \frac{\bar{p} - p_{H_0}}{\frac{\sigma}{\bar{p}}}$$

Where,

$\bar{p}$ = sample proportion

p_{H_0} = the value of p under the assumption that null hypothesis is true

$\frac{\sigma}{\bar{p}}$ = Standard error of sample proportion

The value of $\frac{\sigma}{\bar{p}}$ is computed by using the following formula:

$$\frac{\sigma}{\bar{p}} = \frac{\sqrt{p_{H_0} q_{H_0}}}{n}$$

Where, $q_{H_0} = 1 - p_{H_0}$

n = Sample size

For a given level of significance α , the computed value of Z is compared with the corresponding critical values, i.e. $Z_{\alpha/2}$ or $-Z_{\alpha/2}$ to accept or reject the null hypothesis. We will consider a few examples to explain the testing procedure for a single population proportion.

Example 8.8: An officer of the health department claims that 60 per cent of the male population of a village comprises smokers. A random sample of 50 males showed that 35 of them were smokers. Are these sample results consistent with the claim of the health officer? Use a level of significance of 0.05.

Solution:

Sample size (n) = 50

Sample proportion = $\bar{p} = \frac{x}{n} = \frac{35}{50} = 0.70$

$H_0 : p = 0.60$

$H_1 : p > 0.60$

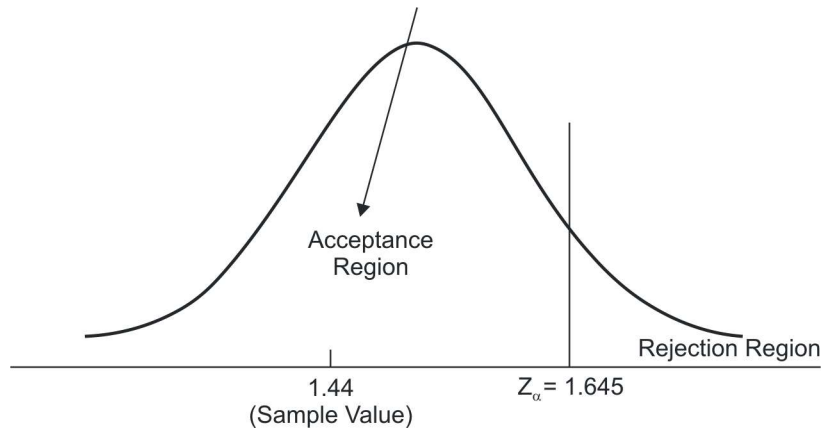
The test statistic is given by:

$$Z = \frac{\bar{p} - p_{H_0}}{\frac{\sigma}{\bar{p}}} = \frac{0.70 - 0.60}{0.069} = \frac{0.10}{0.069} = 1.44$$

NOTES

$$\left(\frac{\sigma}{p} = \sqrt{\frac{P_{H_0} q_{H_0}}{n}} = \sqrt{\frac{0.6 \times 0.4}{50}} = \sqrt{\frac{0.24}{50}} = 0.069 \right)$$

It is a one-tailed test. For a given level of significance $\alpha = 0.05$, the critical value of Z is given by $Z_\alpha = Z_{0.05} = 1.645$. It is seen that the sample value of $Z = 1.44$ lies in the acceptance region as shown below (see figure).



Therefore, there is not enough evidence to reject the null hypothesis. So it can be concluded that the proportion of male smokers is not statistically different from 0.60.

8.6.4 Tests Significance of Difference between two Population Proportions

Here, the interest is to test whether the two population proportions are equal or not. The hypothesis under investigation is:

$$H_0 : p_1 = p_2 \rightarrow p_1 - p_2 = 0$$

$$H_1 : p_1 \neq p_2 \rightarrow p_1 - p_2 \neq 0$$

The alternative hypothesis assumed is two sided. It could as well have been one sided. The test statistic is given by:

$$Z = \frac{\bar{p}_1 - \bar{p}_2 - (p_1 - p_2)H_0}{\sigma_{\bar{p}_1 - \bar{p}_2}}$$

Where,

$\bar{p}_1$ = Sample proportion possessing a particular attribute from population 1

$\bar{p}_2$ = Sample proportion possessing a particular attribute from population 2

$\sigma_{\bar{p}_1 - \bar{p}_2}$ = Standard error of difference between proportions.

NOTES

$(p_1 - p_2)_{H_0}$ = Value of difference between population proportion under the assumption that the null hypothesis is true.

The formula for $\sigma_{\bar{p}_1 - \bar{p}_2}$ is given by:

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

We do not know the value of p_1, p_2 , etc., but under the null hypothesis $p_1 = p_2 = p$.

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{pq}{n_1} + \frac{pq}{n_2}} = \sqrt{pq \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

The best estimate of p is given by:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Where,

x_1 = Number of successes in sample 1

x_2 = Number of successes in sample 2

n_1 = Size of sample taken from population 1

n_2 = Size of sample taken from population 2

It is known that $\bar{p}_1 = \frac{x_1}{n_1}$ and $\bar{p}_2 = \frac{x_2}{n_2}$.

Therefore, $x_1 = n_1 \bar{p}_1$ and $x_2 = n_2 \bar{p}_2$

Therefore, $\hat{p} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2}$

Therefore, the estimate of standard error of difference between the two proportions is given by:

$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\hat{p}\hat{q} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Where $\hat{p}$ is as defined above and $\hat{q} = 1 - \hat{p}$. Now, the test statistic may be rewritten as:

$$Z = \frac{\bar{p}_1 - \bar{p}_2 - (p_1 - p_2)_{H_0}}{\sqrt{\hat{p}\hat{q} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Now, for a given level of significance α , the sample Z value is compared with the critical Z value to accept or reject the null hypothesis. We consider below a few examples to illustrate the testing procedure described above.

NOTES

Example 8.9: A company is interested in considering two different television advertisements for the promotion of a new product. The management believes that advertisement A is more effective than advertisement B. Two test market areas with virtually identical consumer characteristics are selected. Advertisement A is used in one area and advertisement B in the other area. In a random sample of 60 consumers who saw advertisement A, 18 tried the product. In a random sample of 100 customers who saw advertisement B, 22 tried the product. Does this indicate that advertisement A is more effective than advertisement B, if a 5 per cent level of significance is used?

Solution:

$$H_0 : p_a = p_b$$

$$H_1 : p_a > p_b$$

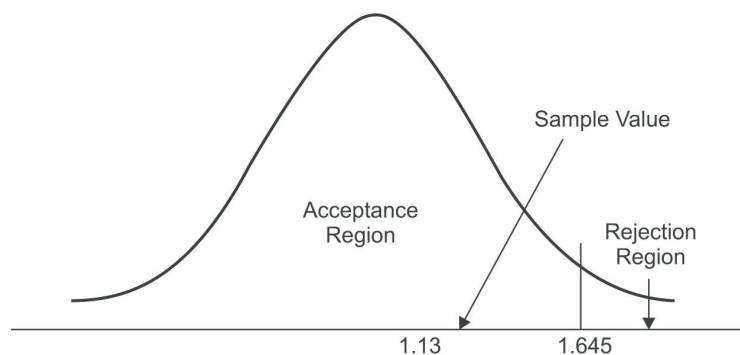
$$n_A = 60, \quad x_A = 18, \quad n_B = 100, \quad x_B = 22$$

$$\left(\bar{p}_A = \frac{x_A}{n_A} = \frac{18}{60} = 0.3 \right) \quad \left(\bar{p}_B = \frac{x_B}{n_B} = \frac{22}{100} = 0.22 \right)$$

$$Z = \frac{\bar{P}_A - \bar{P}_B - (p_A - p_B)H_0}{\frac{\sigma_{\bar{P}_A - \bar{P}_B}}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}} = \frac{0.3 - 0.22 - 0}{\frac{0.08}{\sqrt{0.25 \times 0.75 \left(\frac{1}{60} + \frac{1}{100}\right)}}} = \frac{0.08}{\sqrt{0.25 \times 0.75 (0.0267)}} = \frac{0.08}{0.071} = 1.3$$

$$\left(\hat{p} = \frac{x_A + x_B}{n_A + n_B} = \frac{18 + 22}{60 + 100} = \frac{40}{160} = 0.25 \right)$$

The critical value of Z at 5 per cent level of significance is 1.645. The sample value of Z = 1.13 lies in the acceptance region as shown in the figure below:



Check Your Progress

9. What is necessary for a T-distribution test?
10. What is one interested in knowing in the tests about proportion?

8.7 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

NOTES

1. Descriptive hypothesis concerns the existence, size, form or distribution of some concept subject to verification.
2. Relational hypothesis is proposed in order to test the relationship or linkage between two variables.
3. The two hypotheses are generally referred to as the (1) null hypotheses denoted by H_0 and (2) alternative hypothesis denoted by H_1 .
4. The rejection of the null hypothesis indicates that the differences/relationship have a statistical significance and the acceptance of the null hypothesis means that any difference/relationship is due to chance.
5. A hypothesis may be rejected or accepted depending upon whether the value of the test statistic falls in the rejection or the acceptance region.
6. Hypothesis testing can be broadly divided into two types, which are as follows:
 - Parametric tests or standard tests of hypothesis
 - Non-parametric tests or distribution-free tests of hypothesis
7. Some of the popular parametric tests are z-test t-test and F-test.
8. Z-test is helpful in judging the significance of difference between the means of two independent large samples, to compare the sample proportion to a theoretical value of population proportion and to judge the significance of median, mode and coefficient of correlation.
9. For a T-distribution test, it is necessary that the observations in the two samples should be collected in the form of matched pairs. It means that each observation in one sample must be paired with an observation in the other sample in such a way that they are matched excluding the other factors which do not fall within the area of this test.
10. In the tests about proportion, one is interested in examining whether the respondents possess a particular attribute or not.

8.8 SUMMARY

- In the words of *Lunberg* a hypothesis is “a tentative generalization the validity of which to be tested”. Thus, hypothesis is any guess, imaginative idea which becomes a basis for action or investigation.
- Hypotheses are not given to us readymade. This is true especially in social sciences where there has not yet evolved a highly developed theoretical system in many areas of its subject matter, affording fruitful bases for hypothesis formulation.

- There can be two basic type of hypothesis, viz., (i) descriptive and (ii) relational.
- Correlational relationships state merely that the variables occur together in some specified manner without implying that one causes the other. Such weak claims are often made when we believe there are more basic causal forces that affect both variables or when we have not developed enough evidence to claim a stronger linkage.
- With explanatory, (causal), hypotheses, there is an implication that the existence of, or a change in one variable causes or leads to an effect on the other variable.
- An investigation, in order to be practical, should relate the hypothesis to the investigational techniques feasible in the particular science.
- The null hypothesis is the hypothesis of the population parameter taking a specified value. In case of two populations, the null hypothesis is of no difference or the difference taking a specified value. The hypothesis that is different from the null hypothesis is the alternative hypothesis.
- The rejection of the null hypothesis indicates that the differences/relationship have a statistical significance and the acceptance of the null hypothesis means that any difference/relationship is due to chance.
- Before a sample is drawn from the population, it is very important to specify the values of test statistic that will lead to rejection or acceptance of the null hypothesis. The one that leads to the rejection of null hypothesis is called the critical region.
- The hypothesis may be rejected or accepted depending upon whether the value of the test statistic falls in the rejection or the acceptance region. Management decisions are based upon the statistical decision of either rejecting or accepting the null hypothesis.
- Non-parametric tests have less stringent and fewer assumptions. They do not specify normally distributed populations or equality of variances.
- Parametric tests assume certain properties of the population sample such as observations from a normal population, large sample size, population parameters like mean and variance.
- The T-test is based on t-distribution and is used to judge the significance of a sample mean or the difference between the means of two small samples when the population variance is not known.
- One-sample sign test is a very simple distribution-free test and is applied in case of a sample from a continuous symmetrical population, wherein the probability of a sample to be either less or more than mean is half.
- The method for comparing two related samples in hypothesis testing is the paired t-test. For this test, it is necessary that the observations in the two samples should be collected in the form of matched pairs.

NOTES

NOTES

8.9 KEY WORDS

- **Null Hypothesis:** It is a type of hypothesis that proposes that there is no difference between certain characteristics of a population (or data-generating process).
- **Parametric Tests:** These are tests that make assumptions about the parameters of the population distribution from which the sample is drawn.
- **Non-Parametric Test:** These are tests which do not make assumptions about the frequency distribution of variables that are to be evaluated.
- **T-test:** It is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features.

8.10 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. State two examples of explanatory hypothesis.
2. What does the level of significance in hypothesis testing denote?
3. Write a short note on the determination of critical region.
4. Name some non-parametric tests.
5. Write a short note on the Fisher-Irwin test.

Long-Answer Questions

1. Discuss the ways in which a hypothesis may be developed.
2. Elaborate upon the tests involved in hypothesis testing.
3. Explain the standard tests of hypothesis.

8.11 FURTHER READINGS

Saravanel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.

Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.

Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.

Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.

Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

BLOCK - III
DATA ANALYSIS AND INTERPRETATION

*Data Processing and
Preliminary Data
Analysis*

**UNIT 9 DATA PROCESSING AND
PRELIMINARY DATA
ANALYSIS**

NOTES

Structure

- 9.0 Introduction
- 9.1 Objectives
- 9.2 Survey Field Work and Data Collection
- 9.3 Nature and Scope of Data Preparation: Editing and Coding
 - 9.3.1 Data Entry and Data Cleaning
- 9.4 Preliminary Data Analysis
 - 9.4.1 Assessing for Normality and Outliers
- 9.5 Answers to Check Your Progress Questions
- 9.6 Summary
- 9.7 Key Words
- 9.8 Self Assessment Questions and Exercises
- 9.9 Further Readings

9.0 INTRODUCTION

Data processing includes the collection and manipulation of data in order to produce useful and meaningful information. After the collection of data, it is checked to ensure that there are no errors in the conduct of the study. The process includes data editing and coding wherein the former is concerned with the evaluation of information for any inconsistencies and the latter is concerned with identifying and denoting a numeral to the responses given by respondents. A preliminary data analysis is also carried out in order to organize data into homogenous categories, describe the data and so on. In this unit, the concepts of data collection, editing, coding and preliminary data analysis have been explained.

9.1 OBJECTIVES

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

- Explain the concepts of Survey Field Work and Data Collection
- Discuss the nature and scope of Data Preparation
- Describe Preliminary Data Analysis

9.2 SURVEY FIELD WORK AND DATA COLLECTION

NOTES

The first step in the processing begins post the questionnaire/or primary data survey. The researcher needs to validate the fieldwork to check whether the execution of the study was handled properly. Thus, he must meticulously go over all the raw data forms and check them for errors and find out whether in the conducted interviews or schedules a standardized set of instructions and reporting was followed or not. As we stated earlier in Unit 5, considerable validation is done at the pilot testing stage of the questionnaire formation. The significance of the validation becomes more important in the following cases:

- In case the form had been translated into another language, expert analysis to see whether the meaning of the questions in the two measures is the same or not. The second validation is done by measuring the reliability index of the original and the translated form.
- The second case could be that the questionnaire survey has to be done at multiple locations and one has outsourced to an outside research agency. In this case, it might be essential to carry out checks during the fieldwork as well to ensure that the process being followed is correct. As here there is both a time and a cost element involved, in case the investigators are erring it needs to be corrected immediately.

Post the survey there might be instances when the survey questionnaire cannot be used for analysis for multiple causes. It might be that:

- The answers that have been obtained and the question instructions that were given, such as qualifying instructions like, 'in case answer is _____ please answer the next set of questions, else go to question _____.' Were completely overlooked.
- The respondent seems to have used the same response category for all the questions; for example there is a tendency on a five point scale to give 3 as the answer for all questions.
- The form that is received back is incomplete, in the sense that either the person has not filled the answer to all questions, especially the open-ended ones, or in case of a multiple-page questionnaire, one or more pages are missing.
- The questionnaire is filled by someone who is not a representative of the population under study. For example, in a study on two-wheeler owners perception of Tata small car, Nano, people who have either no vehicle currently or have a small car might have filled in the questionnaire.
- The filled-in form is received after the deadline for receiving the questionnaires has elapsed and the researcher is on the data analysis and interpretation stage.

- The forms received are not in the proportion of the sampling plan. For example, instead of an equal representation from government and private sector employees, 65 per cent of the forms are from the government sector. In such a case the researcher either would need to discard the extra forms or get an equal number filled-in from private sector employees.

NOTES

Check Your Progress

1. When does the first step in processing begin?
2. How is second validation done?

9.3 NATURE AND SCOPE OF DATA PREPARATION: EDITING AND CODING

Data preparation refers to the process of preparing the raw data to make it suitable for the actual data processing and analysis. This process comes before the data processing step and so is involved in activities like data editing, coding, entry and cleaning. The nature of these activities is to make corrections in the raw data, to form logical sets, reformatting data where required. Activities like proofing, cataloguing, structuring, enrichment and curation all come under the scope of data preparation. This is important for instilling confidence in the data provided to the researcher, it saves time and effort and helps the research avoid unnecessary or superfluous data collected.

Data editing is the process that involves detecting and correcting errors (logical inconsistencies) in data. After collection, the data is subjected to processing. Processing requires that the researcher must go over all the raw data forms and check them for errors.

Once the validation process has been completed, the next step is the editing of the raw data obtained. While carrying out editing, the researcher needs to ensure that:

- The data obtained is complete in all respects.
- It is accurate in terms of information recorded and responses sought.
- Questionnaires are legible and are correctly deciphered, especially the open-ended questions.
- The response format is in the form that was instructed.
- The data is structured in a manner that entering the information will not be a problem.

The editing process is carried out at two levels, the first of these is field editing and the second is central editing.

NOTES

(i) Field Editing

Usually, the preliminary editing of the information obtained is done by the field investigators or supervisors who review the filled forms for any inconsistencies, non-response, illegible responses or incomplete questionnaires. Thus the errors can be corrected immediately and if need be the respondent who filled in the form, can be contacted again. The other advantage is that regular field editing ensures that one can also check that the surveyor is able to handle the process of instructions and probing correctly or not. Thus, the researcher can advise and train the investigator on how to administer the questionnaire correctly.

(ii) Centralized in-house Editing

The second level of editing takes place at the researcher's end. At this stage, there are two kinds of typical problems that the researcher might encounter.

First, one might detect an incorrect entry. For example, in case of a five-point scale one might find that someone has used a value more than 5. In another case, one might be asking a question like, 'how many days do you travel out of the city in a week?' and the person says '15 days'. Here one can carry out a quick frequency check of the responses; this will immediately detect an unexpected value.

The second and the major problem that most researchers face is that of 'armchair interviewing' or a fudged interview. One way to handle this is to first scroll the answers to the open-ended questions, as generally if the investigator is filling in multiple forms faking these would be difficult.

The researcher has some standard processes available to him to carry out the editing process. These are briefly discussed below.

- **Backtracking:** The best and the most efficient way of handling unsatisfactory responses is to return to the field, and go back to the respondents. This technique is best used for industrial surveys but it is a little difficult in individual surveys.
- **Allocating missing values:** This is a contingency plan that the researcher might need to adopt in case going back to the field is not possible. Then the option might be to assign a missing value to the blanks or the unsatisfactory responses. However, this works in case:
 - o The number of blank or wrong answers is small.
 - o The number of such responses per person is small.
 - o The important parameters being studied do not have too many blanks, otherwise the sample size for those variables becomes too small for generalizations.
- **Plug value:** In cases such as the third condition above, when the variable being studied is the key variable, then sometimes the researcher might insert a *plug value*. Sometimes one can plug an average or a neutral value in such

cases, for example a 3 for a five-point scale or the researcher might have to establish a rule as to what value will be put if the person has not answered. Sometimes, the respondents' pattern of responses to other questions is used to extrapolate and calculate an appropriate response for the missing answer.

- **Discarding unsatisfactory responses:** If the response sheet has too many blanks/illegible or multiple responses for a single answer, the form is not worth correcting and editing. Hence, it is much better to completely discard the whole questionnaire.

Coding

The process of identifying and denoting a numeral to the responses given by a respondent is called *coding*. This is essentially done in order to help the researcher's in recording the data in a tabular form later. It is advisable to assign a numeric code even for the categorical data (e.g., gender). In fact, even for open-ended questions, which are in a statement form, we will try to categorize them into numbers. The reason for doing this is that the graphic representation of data into charts and figures becomes easier.

Usually, the codes that have been formulated are organized into fields, records and files. For example, the gender of a person is one *field* and the codes used could be 0 for males and 1 for females. All related fields, for example, all the demographic variables like age, gender, income, marital status and education could be one *record*. The records of the entire sample under study form a single *file*. The data that is entered in the spreadsheet, such as on EXCEL, is in the form of a data matrix, which is simply a rectangular arrangement of the data in rows and columns. Here, every row represents a single case or record. For example, consider the following representation from a study on two-wheeler buyers (Table 9.1):

Table 9.1 Sample Record: Excel Sheet for Two-wheeler Owners

Unit Column 1	Occupation Column 2	Vehicle Column 3	Km/day Column 4	Marital status Column 5	Family size Column 6
1	4	1	20	1	3
2	3	2	25	2	1
3	5	1	25	1	4
4	2	1	15	2	2
5	4	2	20	2	4

Here, the data matrix reveals that each field is denoted on the column head and each case record is to be read along the row. The data in the first column represents the unique identification given to a particular respondent (also marked on his/her questionnaire). The second column has data entered on the basis of a coding scheme where every occupation is given a number value (for example, 1 stands for government service and 5 stands for student and so on). Column 3 has 1 representing a motorcycle and 2 representing a scooter. The next value is of the average number of kilometres a person travels per day.

NOTES

NOTES

This is followed by the marital status, with 1 signifying unmarried and 2 married. The last column is again a ratio scale data with the number of family members. The researcher can enter the data on the spreadsheet of the software package he/she is using for the analysis.

Codebook formulation: In order to manage the data entry process, it is best to prepare a method for entering the records. This coding scheme for all the variables under study is called a *code book*. Generally, while designing the rules, care must be taken to decide on some categories that are:

- **Comprehensive:** Should cover all the possible answer to the question that was asked.
- **Mutually exclusive:** The categories and codes devised must be exclusive or clearly different from each other.
- **Single variable entry:** The response that is being entered and the code for it should indicate only a single variable. For example, a 'working single mother' might seem an apparently simple category which one could code as 'occupation'. However, it needs three columns—occupation, marital status and family life cycle. So, one needs to have three different codes to enter this information.

Based on the above rules, one creates a code book. This would generally contain information on the question number, variable name, response descriptors and coding instructions and the column descriptor.

As we have read in Unit 6, a questionnaire can have both closed-ended and open-ended questions. When the questions are structured and the response categories are prescribed then one does what is called *pre-coding*, i.e., giving numeral codes to the designed responses *before* administration. However, if the questions are structured and the answers are open ended, one needs to decide on the codes *after* the administration of the survey. This is called *post-coding*.

Coding Closed-ended Structured Questions

The method of coding for structured questions is easier as the response categories are decided in advance. The coding method to be followed for different kinds of questions is discussed below.

Dichotomous questions: For dichotomous questions, which are on a nominal scale, the responses can be binary, for example:

Do you eat ready-to-eat food? Yes = 1; no = 0.

This means if someone eats ready- to- eat food he/she will be given a score of 1 and if not, then 0.

Table 9.2 Codebook Extract for Ready-to-eat Food Study

Data Processing and
Preliminary Data
Analysis

Question No.	Variable Name	Coding Instruction	Symbol used for Variable Name
1.	Buy ready-to-eat food products	Yes = 1 No = 0	X1
2.	Use ready-to-eat food products	Yes = 1 No = 0	X2
22.	Age	Less than 20 years = 1, 21–26 years = 2, 27–35 years = 3, 36–45 years = 4, More than 45 years = 5	X22
23.	Gender	Male = 1 Female = 2	X23
24.	Marital status	Single = 1 Married = 2 Divorced/widow = 3	X24
25.	No. of children	Exact no. to be written	X25

NOTES

Ranking questions: For ranking questions where there are multiple objects to be ranked, the person will have to make multiple columns, with column numbers equaling the number of objects to be ranked. For example, for ranking chocolates, the code book would be as follows:

Q.No.	Variable Name	Coding Instructions	Variable Name
1.	<i>Cadbury Silk</i>	Number from 1-10	X 10a
2.	<i>Snickers</i>	Number from 1-10	X 10b
3.	<i>Bournville</i>	Number from 1-10	X 10c
4.	<i>Gems</i>	Number from 1-10	X 10d
5.	<i>Ferrero Rocher</i>	Number from 1-10	X 10e

Checklists/multiple responses: In questions that permit a large number of responses, each possible response option should be assigned a separate column. For example, consider the following question:

Which of the following newspapers do you read? (Tick all that you read.)

Times of India	_____
Hindustan Times	_____
Mail Today	_____
Indian Express	_____
Deccan Chronicle	_____
Asian Age	_____
Mint	_____

NOTES

For this question, the number of columns required is seven, one for each newspaper. The coding instructions for each column would be as follows: in case the person ticks on a name, the paper = 1, and in case he does not tick, the paper = 0.

Scaled questions: For questions that are on a scale, usually an interval scale, the question/statement will have a single column and the coding instruction would indicate what number needs to be allocated for the response options given in the scale. Consider the following questions.

Please indicate level of your agreement with the following statements.

	Compared to the Past (5-10 years)	SA	A	N	D	SD
1	The individual customer today shops more					
2	The consumer is well informed about market offerings					

SA – Strongly agree; A – Agree; N – Neutral; D – Disagree; SD – Strongly disagree

The code book for this will look as follows:

Col.no.	Variable Name	Coding Instructions	Variable Name
1.	Individual shops more	A number from 1–5 SA = 5, A = 4, N = 3, D = 2, SD = 1	X 1a
2.	Well informed	- do -	X 1b

Missing values: It is advisable to use a standard format for signifying a non-response or a missing value. For example, a code of 9 could be used for a single-column variable, 99 for a double-column variable, and 999 for a three character variable and so on. The researcher must take care as far as possible to use a value that is starkly different from the valid responses. This is one of the reasons why 9 is suggested. However, in case you have a 10 point scale do not use 9.

Coding Open-ended Structured Questions

The coding of open-ended questions is quite difficult as the respondents' exact answers are noted on the questionnaire. Then the researcher (either individually or as a team) looks for patterns and assigns a category code.

The following example is an open ended question:

If you think lean management was a success so far, please specify three most significant reasons that have contributed to its success in your opinion.

People gave different answers. Thus, based upon the responses obtained, for the above question, the following post-code book was created:

NOTES

Col.No.	Variable Name	Coding Instructions	Variable Name
63	Improvement at workplace by eliminating waste.	Yes = 1 No = 0	X 63a
64	To meet increasing demands of customers	Yes = 1 No = 0	X 63b
65	To improve quality	Yes = 1 No = 0	X 63c
66	To achieve corporate goal	Yes = 1 No = 0	X 63d

9.3.1 Data Entry and Data Cleaning

Let us discuss data entry and data cleaning here. Data cleaning refers to the process of removing or modifying duplicated, incomplete, irrelevant or duplicated data. Data entry as the name suggest refers to the logical entry of data.

Sometimes, the data obtained from the primary instrument is so huge that it becomes difficult to interpret. In such cases, the researcher might decide to reduce the information into homogenous categories. This method of arrangement is called classification of data. This can be done on the basis of class intervals.

Classification by class intervals: Numerical data, like the ratio scale data, can be classified into *class intervals*. This is to assist the quantitative analysis of data. For example, the age data obtained from the sample could be reduced to homogenous grouped data, for example all those below 25 form one group, those 25–35 are another group and so on. Thus, each group will have class limits—an upper and a lower limit. The difference between the limits is termed as the class magnitude. One can have class intervals of both equal and unequal magnitude.

The decision on how many classes and whether equal or unequal depends upon the judgement of the researcher. Generally, multiples of 2 or 5 are preferred. Some researchers adopt the following formula for determining the number of class intervals:

$$I = R / (1 + 3.3 \log N)$$

where,

I = size of class interval,

R = Range (i.e., difference between the values of the largest item and smallest item among the given items),

N = Number of items to be grouped.

The class intervals that are decided upon could be *exclusive*, for example:

10–15

15–20

20–25

25–30

NOTES

In this case, the upper limit of each is excluded from the category. Thus we read the first interval above as 10 and under 15, the next one as 15 and under 20 and so on.

The other kind is *inclusive*, that is:

10–15

16–20

21–25

26–30

Here, both the lower and the upper limits are included in the interval. It says 10–15 but actually means 10–15.99. It is recommended that when one has continuous data it should be signified as 10–15.99, as then all possibilities of the responses are exhausted here. However, for discrete data one can use 10–15.

Once the categories and codes have been decided upon, the researcher needs to arrange the same according to some logical pattern. This is referred to as **tabulation** of data. This involves an orderly arrangement of data into an array that is suitable for a statistical analysis. Usually, this is an orderly arrangement of the rows and columns. In case there is data to be entered for one variable, the process is a simple tabulation and, when it is two or more variables, then one carries out a cross-tabulation of data.

9.4 PRELIMINARY DATA ANALYSIS

Once the data has been cleaned and entered in a tabular form, the researcher is advised to do a preliminary data exploration, in order to assess the expected trends of the findings. Sometimes, these indicative trends may demonstrate that the data collection or instrument design is faulty and needs some corrections. Thus, before one goes about testing the formulated hypotheses, one carries out a loosely structured exploration. Most of the exploration is done on the basis of the graphical and visual display of the data patterns that seem to be emerging. In this section we will discuss some widely used and simplistic measures of displaying data.

- **Bar and pie charts:** The data that is available as classification or demographic variable is most often on a categorical or nominal scale. Thus, the tabled data can be plotted to demonstrate the pattern of responses.
- **Histogram:** For metric–interval and ratio scale data, the data is represented through a histogram. The representation would be able to demonstrate the distribution pattern in terms of whether it is normally distributed or demonstrates skewness.
- **Stem and leaf displays:** This is another way of displaying the metric data. It is very easy to compute and can be done manually or with the help of Minitab. It shows individual data values in each set as against the histogram which presents only group aggregates. It shows the pattern of responses in

each interval and yet can maintain the rank order for a quick approximation of the median or quartile. Each row or line is called a stem and each value on the line is a leaf.

There are other methods like box plots, which are a more detailed representation as compared to histograms. These are basically descriptive statistical values for the data obtained and these are based upon the measures of central tendency and dispersion. These will be discussed in further units.

9.4.1 Assessing for Normality and Outliers

Let us begin with assessing for normality of data.

Normality

An assessment of the normality of data is a requirement for many statistical tests because normal data is an underlying assumption in parametric testing. There are two main methods of assessing normality: Graphical and numerical (including statistical tests). Statistical tests have the advantage of making an objective judgment of normality but have the disadvantage of sometimes not being sensitive enough at low sample sizes or overly sensitive to large sample sizes. Graphical interpretation has the advantage of allowing good judgment to assess normality in situations when numerical tests might be over or undersensitive. Although normality assessment using graphical methods need a great deal of the experience to avoid the wrong interpretations. If we do not have a good experience, it is the best to rely on the numerical methods. There are various methods available to test the normality of the continuous data, out of them, most popular methods are Shapiro–Wilk test, Kolmogorov–Smirnov test, skewness, kurtosis, histogram, box plot, P–P Plot, Q–Q Plot, and mean with SD. The two well-known tests of normality, namely, the Kolmogorov–Smirnov test and the Shapiro–Wilk test are most widely used methods to test the normality of the data.

Outliers

Outliers are data points that are far from other data points. In other words, they're unusual values in a dataset. Outliers are problematic for many statistical analyses because they can cause tests to either miss significant findings or distort real results. There are no strict statistical rules for definitively identifying outliers. Finding outliers depends on subject-area knowledge and an understanding of the data collection process.

Check Your Progress

3. List two points that a researcher needs to ensure while carrying out editing.
4. What is a code book?
5. What is tabulation of data?
6. What is the advantage and disadvantage of statistical tests?

NOTES

NOTES

9.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The first step in the processing begins post the questionnaire/or primary data survey.
2. The second validation is done by measuring the reliability index of the original and the translated form.
3. While carrying out the editing, the researcher needs to ensure that:
 - The data obtained is complete in all respects.
 - It is accurate in terms of information recorded and responses sought.
4. In order to manage the data entry process, it is best to prepare a method for entering the records. This coding scheme for all the variables under study is called a *code book*.
5. Once the categories and codes have been decided upon, the researcher needs to arrange the same according to some logical pattern. This is referred to as *tabulation* of data.
6. Statistical tests have the advantage of making an objective judgment of normality but have the disadvantage of sometimes not being sensitive enough at low sample sizes or overly sensitive to large sample sizes.

9.6 SUMMARY

- The first step in the processing begins post the questionnaire/or primary data survey. The researcher needs to validate the fieldwork to check whether the execution of the study was handled properly.
- In case the form had been translated into another language, expert analysis to see whether the meaning of the questions in the two measures is the same or not.
- After collection, the data is subjected to processing. Processing requires that the researcher must go over all the raw data forms and check them for errors.
- Once the validation process has been completed, the next step is the editing of the raw data obtained.
- Usually, the preliminary editing of the information obtained is done by the field investigators or supervisors who review the filled forms for any inconsistencies, non-response, illegible responses or incomplete questionnaires.
- The second level of editing takes place at the researcher's end. At this stage there are two kinds of typical problems that the researcher might encounter.

First, one might detect an incorrect entry. The second and the major problem that most researchers face is that of ‘armchair interviewing’ or a fudged interview.

- The best and the most efficient way of handling unsatisfactory responses is to return to the field, and go back to the respondents. This technique is best used for industrial surveys but a little difficult in individual surveys.
- The process of identifying and denoting a numeral to the responses given by a respondent is called *coding*. This is essentially done in order to help the researcher’s in recording the data in a tabular form later.
- Usually, the codes that have been formulated are organized into fields, records and files. For example, the gender of a person is one *field* and the codes used could be 0 for males and 1 for females.
- In order to manage the data entry process, it is best to prepare a method for entering the records. This coding scheme for all the variables under study is called a *code book*.
- The method of coding for structured questions is easier as the response categories are decided in advance.
- For questions that are on a scale, usually an interval scale, the question/ statement will have a single column and the coding instruction would indicate what number needs to be allocated for the response options given in the scale.
- The coding of open-ended questions is quite difficult as the respondents’ exact answers are noted on the questionnaire. Then the researcher (either individually or as a team) looks for patterns and assigns a category code.
- An assessment of the normality of data is a requirement for many statistical tests because normal data is an underlying assumption in parametric testing. There are two main methods of assessing normality: Graphical and numerical (including statistical tests).
- Outliers are data points that are far from other data points. In other words, they’re unusual values in a dataset. Outliers are problematic for many statistical analyses because they can cause tests to either miss significant findings or distort real results.

NOTES

9.7 KEY WORDS

- **Data Editing:** It is the process that involves detecting and correcting errors (logical inconsistencies) in data.
- **Coding:** It is the process of identifying and denoting a numeral to the responses given by a respondent.

- **Data Cleaning:** It refers to a several-step process used to ensure that survey data have been entered correctly both by the interviewer and by data entry personnel.

NOTES

9.8 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. List some advantages of field editing.
2. Write a short note on the problems that occur at the second stage of data editing.
3. List the cases in which allocation of missing values works.
4. Why is the method of coding easier for structured questions?
5. What is the advantage of graphical interpretation?

Long-Answer Questions

1. Discuss some cases in which the significance of validation becomes more important.
2. Analyze the process of data entry and data cleaning.

9.9 FURTHER READINGS

Saravanel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.

Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.

Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.

Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.

Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

UNIT 10 DATA ANALYSIS I

Structure

- 10.0 Introduction
- 10.1 Objectives
- 10.2 Statistics in Data Analysis
- 10.3 Measures of Central Tendency
- 10.4 Measures of Dispersion
- 10.5 Univariate, Bivariate and Multivariate Analysis of Data
- 10.6 Answers to Check Your Progress Questions
- 10.7 Summary
- 10.8 Key Words
- 10.9 Self Assessment Questions and Exercises
- 10.10 Further Readings

NOTES

10.0 INTRODUCTION

Summary statistics is a measure in data analysis that summarizes the collected data or observations with an aim to represent the largest set of data in a simple and lucid manner. The measures included in this are measures of central tendency which include mean, median and mode and measures of dispersion which include standard deviation. Moreover, univariate, multivariate and bivariate analysis of data is also carried out. This unit provides an analysis of the aforementioned concepts.

10.1 OBJECTIVES

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

- Discuss statistics in Data Analysis
- Analyse the measures of Central Tendency
- Explain the measures of Dispersion
- Discuss Univariate, Bivariate and Multivariate Analysis of Data

10.2 STATISTICS IN DATA ANALYSIS

Summary statistics or ‘summary numbers’ are certain measures which are embodiments of some characteristics and which researchers frequently use to typify the data and to briefly describe its important features. The most commonly used summary statistics can be divided into three groups as follows:

1. Percentages
2. Measures of central tendency
3. Measures of dispersion

NOTES

Summary statistics are very highly useful to researchers, both in analyzing the data and in reporting the study findings. If these measures reflect the sample characteristics well, then they become dependable substitutes for the more detailed and unwieldy data obtained from the sample. They are also a useful means of comparing two samples or two sets of data and describing the relationships between them. It is absolutely essential that the marketing researcher understands the summary statistics fully well.

10.3 MEASURES OF CENTRAL TENDENCY

There are three common measures of central tendency. They are as follows:

1. *Mode* : the value of the item that occurs most frequently.
2. *Median* : the value of the middlemost item when all items are arranged in descending or ascending order.
3. *Mean* : the sum of values of all the items divided by the number of items, also known as 'average' or 'arithmetic mean'.

Let us take a simple example in which the values of 200 items are grouped and presented in categories (i.e. classes) of equal width.

Table 10.1

Given		Calculated		
Value (Class) (Rs 4000)	Frequency = <i>f</i> No. of items having value in that class	Mid-value = Midpoint of the class = <i>x</i>	Cumulative frequency (CF)	<i>fx</i>
0-10	6	5	6	30
10-20	49	15	55	735
20-30	88	25	143	2200
30-40	53	35	196	1855
40-50	4	45	200	180
Total	N=200		Total	5000

The approximate values of the mode, median and mean could then be calculated as follows:

Mode: Since the class 20-30 occurs 88 times, which is the maximum number of times, we can take the midpoint of that class viz., 25, as the approximate modal value.

Note: the exact modal value is:

$$20 + \frac{(88 - 49)}{(2 \times 88 - 49 - 53)} \times 10 = 25.3$$

Median: From the CF column, we find that the middlemost item i.e. the 100th item falls in the class 20-30. Therefore, we take the approximate median value as the midpoint of that class, which is again 25.

Note: the exact median value is:

$$20 + \frac{200 - 55}{88} 2 \times 10 = 25.1$$

Mean: We calculate the arithmetic mean ($\bar{x}$) as follows:

$$\bar{x} = \frac{1}{N} \sum fx = \frac{5000}{200} = 25.0$$

What does this 25 mean? If the values, which are in thousands of rupees, are of 'income per month', and if the frequencies are the number of persons having income falling in that class, then this would mean that the average income of the group of 200 persons is Rs. 25,000 p.m.

It may be noted that in this example, all the three measures of central tendency, the mode, median and mean are very close to one another. This need not be the case in all distributions, even though all the three are measures of 'central tendency'. However, in frequency distributions which are nearly symmetric, these three measures may have values close to one another, or even exactly equal values.

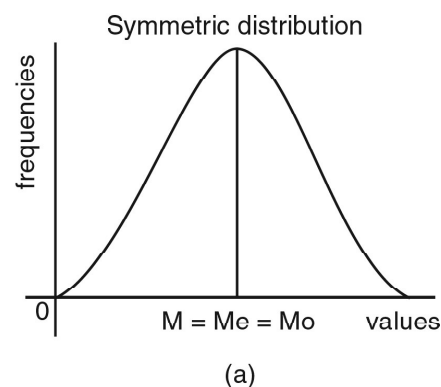
If frequency distributions are presented as graphs, they reveal the nature of the distribution at a glance. Graphs of symmetric and nonsymmetric distributions are shown in the figure 10.1.

The empirical relationship between the three measures of central tendency is as follows:

$$(\text{Mode} - \text{Mean}) = (\text{Median} - \text{Mean})$$

All summary statistics suffer from certain deficiencies when used to describe large bodies of data. The measures of central tendency are no exception. Measures of central tendency could be easily misrepresented or misunderstood if not explained well. The mean, median and mode

NOTES



NOTES

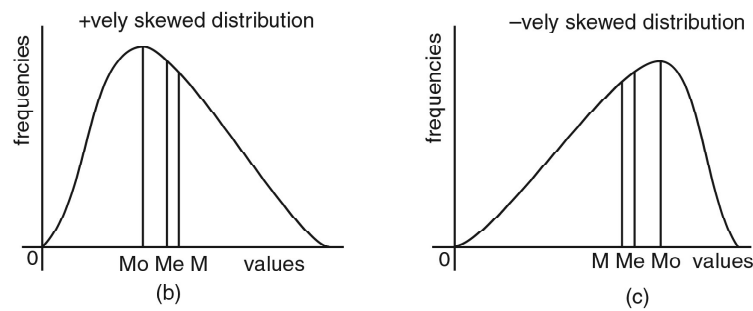


Fig. 10.1

each have their own merits, demerits and exact meanings. Between these three measures, the mean is often preferred because of its suitability for further statistical treatment.

In the above example of 'income per month in thousand Rs', the mean 25 (i.e., Rs. 25,000) is the arithmetic average of the values of the 200 items. Now if these 200 items or members in the sample have been randomly selected from some population then we can generalize and take the sample mean to represent an unbiased estimate of the population mean.

10.4 MEASURES OF DISPERSION

Measures of central tendency locate the centre of the distribution. The study of the spread of a distribution variable can be done through measures of dispersion.

In the above example of sample size 200, the values range from 0 to 50. This range is a simple but crude measure to describe the dispersion or the variation of the values.

The popular measure of dispersion, which is also most suitable for statistical treatment, is *Standard deviation*. The formula for standard deviation (S.D.) is as follows:

$$\text{S.D.} = \left(\frac{1}{N} \sum f(x - \bar{x})^2 \right)^{1/2}$$

Thus this is square root of the mean of the squares of deviations of the values from their mean.

To use this formula, let us make the following calculations from the example given above:

f	$(x - \bar{x})$	$(x - \bar{x})^2$	$f(x - \bar{x})$
6	5-25 = -20	400	2400
49	15-25 = -10	100	4900
88	25-25 = 0	0	0
53	35-25 = +10	100	5300
4	45-25 = +20	400	1600
Total : N = 200			Total : 14,200

Hence, the distribution which has a mean of 25 has S.D. of 8.4.

Note that the range 50 is about 6 times the standard deviation ($8.4 \times 6 = 50.4$). Though this need not always be six times, in normal distributions or distributions close to normal, this is usually the case.

The higher the value of the standard deviation, the higher is the variation or dispersion or spread of the values in the distribution from their mean. This means that they are more scattered from the mean.

The lower the value of S.D., the lower is the spread or scatteredness of the values from their mean. This means that the values in the distribution are more compact and concentrated around the mean.

Correlation and Regression: Correlation is a statistical measure for finding out the degree or strength of association between two or more variables to change together. If two variables x and y are so related that movements in one tend to be accompanied by movements in the other, then x and y are said to be correlated.

The movements of the two variables may be in the same direction or in opposite directions. If x and y move in the same direction they are *positively correlated*. If they move in opposite directions they are *negatively correlated*. For example 'advertising expenditure' and 'sales' may be positively correlated but 'price' and 'sales' may be negatively correlated. If there is no relationship between movements of the two variables, they are not associated with each other and are said to be uncorrelated. The existence of correlation between two variables is no proof that any one of them causes the other. The cause may lie somewhere else. Correlation only proves that the two variables move in sympathy with each other, for whatever reason it may be.

One method of measuring correlation is by using *Karl Pearson's* correlation coefficient:

Here n is the number of pairs of observation and ox and oy are S.Ds of the two variables x and y . The value of this coefficient lies between -1 and $+1$, and it is a pure number which measures the strength of association between the two variables and states whether it is positive or negative. If $r = 0$, the variables are uncorrelated.

Another method of measuring correlation, called *Rank Correlation* is used when the variables cannot be measured but can only be ranked. Rank correlation has a different formula and it is based on the rank or the order, and not on the value of the variable.

Concerned with correlation is another very important concept termed *regression*. Regression is the measure of the average relationship between two or more variables. This relationship is used to estimate or predict the most likely values of one variable for specified values of the other variable. Such relationships are called *regression equations*.

NOTES

10.5 UNIVARIATE, BIVARIATE AND MULTIVARIATE ANALYSIS OF DATA

NOTES

Once the raw data is collected from both primary and secondary sources, the next step is to analyze the same so as to draw logical inferences from them. The data collected in a survey could be voluminous in nature, depending upon the size of the sample. In a typical research study, there may be a large number of variables that the researcher needs to analyze. The analysis could be univariate, bivariate and multivariate in nature. In the univariate analysis, one variable is analyzed at a time. In the bivariate analysis, two variables are analyzed together and examined for any possible association between them. In the multivariate analysis, the concern is to analyse more than two variables at a time. These analyses of data will be discussed in the following unit.

Check Your Progress

1. Name the three common measures of central tendency.
2. Why is mean preferred over median and mode?
3. How are variables analyzed in univariate, bivariate and multivariate analyses?

10.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The three common measures of central tendency are mean, median and mode.
2. Mean is often preferred over median and mode because of its suitability for further statistical treatment.
3. In the univariate analysis, one variable is analysed at a time. In the bivariate analysis two variables are analysed together and examined for any possible association between them. In the multivariate analysis, the concern is to analyse more than two variables at a time.

10.7 SUMMARY

- The most commonly used summary statistics can be divided into three groups as follows:
 - (i) Percentages
 - (ii) Measures of central tendency
 - (iii) Measures of dispersion

- Mean refers to the sum of values of all the items divided by the number of items, also known as '*average*' or '*arithmetic mean*'.
- The empirical relationship between the three measures of central tendency is as follows:

$$(\text{Mode} - \text{Mean}) = (\text{Median} - \text{Mean})$$
- The popular measure of dispersion, which is also most suitable for statistical treatment, is *Standard deviation*.
- The higher the value of the standard deviation, the higher is the variation or dispersion or spread of the values in the distribution from their mean. This means that they are more scattered from the mean.
- If two variables x and y are so related that movements in one tend to be accompanied by movements in the other, then x and y are said to be correlated.
- If x and y move in the same direction they are *positively correlated*. If they move in opposite directions they are *negatively correlated*.
- *Rank Correlation* is used when the variables cannot be measured but can only be ranked. Rank correlation has a different formula and it is based on the rank or the order, and not on the value of the variable.
- The type of statistical techniques used for analysing univariate and bivariate data depends upon the level of measurements of the questions pertaining to those variables.

NOTES

10.8 KEY WORDS

- **Summary Statistics:** These are certain measures which are embodiments of some characteristics and which researchers frequently use to typify the data and to briefly describe its important features.
- **Correlation:** It is a statistical measure for finding out the degree or strength of association between two or more variables to change together.
- **Regression:** It is the measure of the average relationship between two or more variables. This relationship is used to estimate or predict the most likely values of one variable for specified values of the other variable.

10.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define median.
2. Which measure of dispersion is most suitable for statistical treatment?

Long-Answer Questions

1. Discuss the usefulness of summary statistics.
2. Explain the concepts of correlation and regression.

NOTES

10.10 FURTHER READINGS

Saravanel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.

Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.

Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.

Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.

Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

UNIT 11 DATA ANALYSIS II

Structure

- 11.0 Introduction
- 11.1 Objectives
- 11.2 Descriptive Statistics
- 11.3 Univariate Analysis
- 11.4 Bivariate Analysis
- 11.5 Correlation Analysis
- 11.6 Analysis of Variance
- 11.7 Analysis of Covariance
- 11.8 Multivariate Analysis
- 11.9 Conjoint Analysis
- 11.10 Answers to Check Your Progress Questions
- 11.11 Summary
- 11.12 Key Words
- 11.13 Self Assessment Questions and Exercises
- 11.14 Further Readings

NOTES

11.0 INTRODUCTION

Data Analysis is a process of evaluating and describing collected data through the means of statistical or logical techniques. It is essential to maintain data accuracy and integrity for the smooth functioning of a business. For this purpose, proper research is conducted by applying various systematic techniques. Data analyses could be of two types descriptive or inferential. Descriptive Analysis is a such technique that aims to transform raw data into meaningful and easy to understand information. Data analysis involves study of a large number of variables. On this basis, the data analysis could be Univariate, Bivariate and Multivariate in nature. Correlation Analysis, Conjoint analysis, analyses of Variance and Covariance are sub-types used for interpreting data. This unit focuses on the aforementioned concepts and techniques.

11.1 OBJECTIVES

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

- Analyse the concept of Descriptive Statistics
- Discuss Bivariate, Univariate, Multivariate and Correlation Analysis
- Describe the Analysis of Covariance and Conjoint Analysis

11.2 DESCRIPTIVE STATISTICS

NOTES

Descriptive analysis refers to transformation of raw data into a form that will facilitate easy understanding and interpretation. Descriptive analysis deals with summary measures relating to the sample data. The common ways of summarizing data are by calculating average, range, standard deviation, frequency and percentage distribution. The first thing to do when data analysis is taken up is to describe the sample. For example, the average monthly household expenditure on food items in a town is ₹ 2300. About 25 per cent of households spend more than ₹5000 per month on food; 50 per cent of the households spend less than ₹2800 per month on food. Three out of ten households send their children to government schools and 5 per cent of the households go abroad for holidays. Further, these households have earnings of more than ₹ 2 lakh per month. It is also known that the occupation of the head of the household in a town is 15 per cent in business, 30 per cent in the private sector, 45 per cent in government service and the remaining are occupied in odd jobs. Below is a set of typical questions that are required to be answered under descriptive statistics:

- What is the average income of the sample?
- What is the average age of the sample?
- What is the standard deviation of ages in the sample?
- What is the standard deviation of incomes in the sample?
- What percentage of sample respondents are married?
- What is the median age of the sample respondents?
- Which income group has the highest number of user of product in question in the sample?
- Is there any association between the frequency of purchase of product and income level of the consumers?
- Is the level of job satisfaction related with the age of the employees?
- Which TV channel is viewed by the majority of viewers in the age group 20–30 years?

Types of Descriptive Analysis

The type of descriptive analysis to be carried out depends on the measurement of variables into four forms—nominal, ordinal, interval and ratio. Table 11.1 presents the type of descriptive analysis which is applicable under each form of measurement.

Table 11.1 Descriptive analysis for various levels of measurement

Type of Measurement	Type of Descriptive Analysis
Nominal	Frequency table, Proportion percentages, Mode
Ordinal	Median, Quartiles, Percentiles, Rank order correlation
Interval	Arithmetic mean, Correlation coefficient
Ratio	Index numbers, Geometric mean, Harmonic mean

NOTES

It is assumed that readers are acquainted with the methods of descriptive analysis as the material could be found in any elementary text on descriptive statistics. Here only a brief review of some of the methods is mentioned.

Inferential Analysis

After descriptive analysis has been carried out, the tools of inferential statistics are applied. Under inferential statistics, inferences are drawn on population parameters based on sample results. The researcher tries to generalize the results to the population based on sample results. The analysis is based on probability theory and a necessary condition for carrying out inferential analysis is that the sample should be drawn at random. The following is an illustrative list of questions that are covered under inferential statistics based on the previous example.

- Is the average age of the population significantly different from 35?
- Is the average income of population significantly greater than ‘ 25,000 per month?
- Is the job satisfaction of unskilled workers significantly related with their pay packet?
- Do the users and non-users of a brand vary significantly with respect to age?
- Is the growth in the sales of the company statistically significant?
- Does the advertisement expenditure influences sale significantly?
- Are consumption expenditure and disposable income of households significantly correlated?
- Is the proportion of satisfied workers significantly more for skilled workers than for unskilled works?
- Do urban and rural households differ significantly in terms of average monthly expenditure on food?
- Is the variability in the starting salaries of fresh MBA different with respect to marketing and finance specialization?

We will begin our discussion of the univariate and bivariate descriptive analysis with an example as a base. It is on this example that both the analyses techniques will be described. The details of the example since it is length is provided

in the Appendix 1 at the end of the book. Kindly look at the example first to grasp the concepts of univariate and bivariate analyses.

NOTES

Check Your Progress

1. What are the common ways of summarizing data?
2. What does the type of descriptive analysis to be carried out depend on?

11.3 UNIVARIATE ANALYSIS

The first step under univariate analysis is the preparation of frequency distributions of each variable. The frequency distribution is the counting of responses or observations for each of the categories or codes assigned to a variable.

Analysis of Nominal Scale Data with only One Possible Response

Consider a nominal scale variable—gender of respondents in a survey research.

Table 11.2 shows both the raw frequency and the percentages of responses for each category in case of the variable gender in a sample of 414 respondents based on the example given in the appendix.

Table 11.2 Gender of the respondent

		Frequency	Per cent	Valid Per cent	Cumulative Per cent
Valid	Male	301	72.7	72.7	72.7
	Female	113	27.3	27.3	100.0
	Total	414	100.0	100.0	

This tabulation process can be done by hand, using tally marks. The results indicate that out of a sample of 414 respondents, 301 are male and 113 are female. The raw frequencies are often converted into percentages as they are more meaningful. In the present case, for example, there are 72.7 per cent male and 27.3 per cent female respondents.

Analysis of Nominal Scale Data with Multiple Category Responses

At times, the researcher comes across multiple-category questions, where respondents could choose more than one answer. In such a case, the preparation of frequency table and its interpretation is slightly different. If the question in the research study is multiple category question and the responds are allowed to tick more than one choice, the percentage in such a case may not add up to 100. For example, one may consider the following question:

When accessing the internet at a cyber cafe, tick up to four frequently used applications for which you use the cyber cafe.

1. E-mail
2. Chat

3. Browsing
4. Downloading
5. Shopping
6. Net telephony
7. Business and Commerce (e-commerce)
8. Entertainment
9. Adult sites
10. Astrology and Horoscope
11. Education
12. Any other, please specify.

NOTES

The coding for the variable applications has been in binary form where values one and zero are assigned. If the respondent uses a particular application, the value assigned is 1, otherwise 0. The resulting frequency table for the above-mentioned question is as presented in Table 11.3.

Table 11.3 Frequently used Applications at Cyber Cafe

Sl. No.	Application	Frequencies	Percentage (%)
1	Email	399	94.9
2	Chat	316	76.3
3	Browsing	232	56
4	Downloading	197	47.6
5	Shopping	30	7.2
6	Net telephony	30	7.2
7	E-commerce	51	12.3
8	Entertainment	135	32.6
9	Adult sites	59	14.3
10	Astrology and horoscopes	52	12.6
11	Education	159	38.4
12	Any Other	14	3.4
TOTAL RESPONDENTS		414	

*Total exceeds 100% because of multiplicity of answers.

In Table 11.3 the percentages are computed on the total sample size of 414. If these percentages are added up, they would exceed more than 100 per cent. This is because of multiplicity of answers as respondents were given the chance to choose more than one answer. The interpretation of the table would be based on a sample of 414 and is given as:

- The most used application at a cyber cafe is e-mail. It is seen that 94.9 per cent of the users make use of this.
- The second popular application is chatting, and 76.3 per cent of the sample respondents make use of it.

- Similarly, other applications in order of preference are browsing (56 per cent), downloading (47.6 per cent), education 35.4 per cent), entertainment (32.6 per cent) and so on.

NOTES

Analysis of Ordinal Scaled Questions

There could always be some ordinal-scaled questions in the questionnaire. The question before the researcher is how to tabulate and interpret the responses to such questions. It could be done in two ways as would be shown in the following example. The questions asked of the respondents in such a case could be:

- Rank the following five attributes while choosing a restaurant for dinner. Assign a rank of 1 to the most important, 2 to the next important ... and 5 to the least important.
 - Ambience
 - Food quality
 - Menu variety
 - Service
 - Location

From a sample of 32, the responses obtained are given in Table 11.4. To construct univariate tables out of the given data, one can take up one column at a time from Table 9.4 and prepare the separate frequency tables. For example, distribution of rank assigned to attribute food quality may be considered in Table 11.5.

Table 11.4 *Ranking of Various Attributes while Selecting a Restaurant for Dinner*

Respondent No.	Ambience	Food Quality	Menu Variety	Service	Location
1	3	1	4	2	5
2	5	2	1	4	3
3	1	2	5	3	4
4	3	1	5	2	4
5	2	1	5	3	4
6	1	3	2	4	5
7	3	2	4	1	5
8	1	2	5	3	4
9	4	2	3	5	1
10	4	3	1	2	5
11	2	1	5	3	4
12	5	1	4	3	2
13	3	1	5	4	2
14	4	1	2	5	3
15	3	2	5	1	4
16	1	2	5	4	3
17	3	1	4	2	5
18	5	2	1	3	4
19	2	1	4	3	5
20	3	2	4	5	1

21	4	1	5	2	3
22	3	2	1	4	5
23	5	1	4	3	2
24	3	2	5	1	4
25	5	1	4	3	2
26	2	1	3	5	4
27	3	1	4	2	5
28	3	2	1	4	5
29	3	1	5	2	4
30	4	2	1	3	5
31	2	1	5	3	4
32	3	4	1	2	5

NOTES**Table 11.5** *Distribution of Ranks Assigned to Food Quality*

Rank	Frequency	Per cent
1	16	50.0
2	13	40.6
3	2	6.3
4	1	3.1
5	—	—
Total	32	100.0

It is seen from Table 11.5 that out of 32 respondents, 16 (50 per cent) have assigned rank one, 13 (40.6 per cent) ranked two, 2 (6.3 per cent) ranked three and 1 (3.1 per cent) ranked four to food quality. This shows that food quality is given a lot of importance by the respondents. Similar analysis could be carried out for other attributes.

The other way of preparing a univariate table could be to find distribution of attribute which got various ranks. Table 11.6 indicates the distribution of attributes that received rank one.

Table 11.6 *Distribution of Attributes that Received Rank One*

Attribute	Number	Percentage
Ambience	4	12.50
Food Quality	16	50.00
Menu Variety	7	21.88
Service	3	9.38
Location	2	6.25
Total	32	100

Table 11.6 indicates that 50 per cent of the respondents gave food quality rank one, whereas 21.88 per cent gave menu variety as rank one, followed by ambience that was ranked one by 12.5 per cent of the respondents. Similar analysis could be carried out corresponding to the remaining attributes.

NOTES

The ordinal scale data could also be used for preparing a summarized rank order. For example, data presented in Table 11.4 gives the ranking by 32 respondents on five attributes while choosing a restaurant for dinner. The data given in Table 11.4 can be used to prepare the summarized rank ordering of various attributes. The rankings of attributes given in Table 11.4 can be presented in the form of frequency distribution in Table 11.7.

Table 11.7 Frequency Table of the Rankings of the Attributes while Selecting a Restaurant for Dinner

Attribute	Rank				
	1	2	3	4	5
Ambience	4	5	13	5	5
Food Quality	16	13	2	1	0
Menu Variety	7	2	2	9	12
Service	3	8	11	6	4
Location	2	4	4	11	11
Total	32	32	32	32	32

To calculate a summary rank ordering, the attribute with the first rank was given the lowest number (1) and the least preferred attribute was given the highest number (5).

The summarized rank order is obtained with the following computations as:

Ambience	:	$(4 \times 1) + (5 \times 2) + (13 \times 3) + (5 \times 4) + (5 \times 5)$	=	98
Food Quality	:	$(16 \times 1) + (13 \times 2) + (2 \times 3) + (1 \times 4) + (0 \times 5)$	=	52
Menu Variety	:	$(7 \times 1) + (2 \times 2) + (2 \times 3) + (9 \times 4) + (12 \times 5)$	=	113
Service	:	$(3 \times 1) + (8 \times 2) + (11 \times 3) + (6 \times 4) + (4 \times 5)$	=	96
Location	:	$(2 \times 1) + (4 \times 2) + (4 \times 3) + (11 \times 4) + (11 \times 5)$	=	121

The total lowest score indicates the first preference ranking. The results show the following rank ordering:

- (1) Food quality
- (2) Service
- (3) Ambience
- (4) Menu variety
- (5) Location

11.4 BIVARIATE ANALYSIS

As already mentioned, bivariate analysis examines the relationship between two variables. There are various methods used for carrying out bivariate analysis. We will discuss two methods, namely, cross-tabulation and correlation coefficient.

(i) Cross-Tabulation

In simple tabulation, the frequency and the percentage for each question was calculated. In cross-tabulation, responses to two questions are combined and data is tabulated together. A cross-tabulation counts the number of observations in each cross-category of two variables. The descriptive result of a cross-tabulation is a frequency count for each cell in the analysis. For example, in cross-tabulating a two-category measure of income (low- and high-income households) with a two-category measure of purchase intention of a product (low and high purchase intentions) the basic result is a cross-classification as shown in Table 11.8.

Table 11.8 *Cross-table of Purchase Intention and Income*

		Income	
		Low Income	High Income
Purchase Intention	Low purchase intention	120	60
	High purchase intention	80	190
		200	250

The results of cross-tabulation show the number of sample respondents with low income having low purchase intention, low income with high purchase intention, high income with low purchase intention and high income with high purchase intention.

As is the case with simple tabulations, the results of a cross-tabulation are more meaningful if cell frequencies are computed as percentages. The percentages can be computed in three-ways. As is the case of Table 11.8, the percentages can be computed (i) row-wise so that the percentages in each row add up to 100 per cent; (2) column-wise so that the percentages in each column add up to 100 per cent or (3) cell percentages, such that percentages added across all cells equal 100 per cent. The interpretation of percentages is different in each of the three cases. Therefore, the question arises which of these percentages is most useful to the researcher. What is the general rule for computing percentages?

The basis for calculating category percentage depends upon the nature of relationship between the variables. One of the variables could be viewed as dependent variable and the other one as independent variable. In the cross-tabulation presented in Table 11.8, the purchase intention could be treated as dependent variable, which depends upon income (independent variable). The rule is to cast percentages in the direction of independent (causal) variable across the dependent variable. For Table 11.8, there are 200 respondents with low income, out of which 120 have low purchase intention for the product. In terms of percentages, 60 per cent of the respondents with low income have low purchase intention for the product. Now there are 250 people with high income, out of which 60 have low purchase intention and 190 have high purchase intention for the product. By calculating percentages column wise, it is seen that 24 per cent have low purchase intention whereas 76 per cent have high purchase intention for

NOTES

the product. The results indicate that with increase in income, the purchase intention for the product increases.

Table 11.9 presents the percentages column-wise as given below:

NOTES

Table 11.9 Cross-table of Purchase Intention and Income
(Column-wise Percentages)

		Income	
		Low Income	High Income
Purchase Intention	Low purchase intention	60%	24%
	High purchase intention	40%	76%
		100%	100%

From the above example, it is clear that any two variables, each having certain categories, can be cross-tabulated. The interpretation of the cross-tabulation results may show a high association between two variables. That does not mean one of them, the independent variable, is the cause of the other variable—the dependent variable. Causality between the two variables is more of an assumptions made by the researcher based on his experience or expectations. Just because there is a high association between two variables, it does not imply a cause-and-effect relationship.

(ii) Correlation Coefficient

Simple correlation measures the degree of association between two variables. The correlation could be positive, negative or zero.

Quantitative Estimate of a Linear Correlation

A quantitative estimate of a linear correlation between two variables X and Y is given by Karl Pearson as:

$$r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

which may be rewritten as:

$$r_{xy} = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\sum_{i=1}^n X_i^2 - n\bar{X}^2} \sqrt{\sum_{i=1}^n Y_i^2 - n\bar{Y}^2}}$$

Where, r_{xy} = Correlation coefficient between X and Y

$\bar{X}$ = Mean of the variable X

$\bar{Y}$ = Mean of the variable Y

n = Size of the sample

It may be noted that the above-mentioned formulae are for the linear correlation coefficient. The linear correlation coefficient takes a value between -1

and +1 (both values inclusive). If the value of the correlation coefficient is equal to 1, the two variables are perfectly positively correlated. Similarly, if the correlation coefficient between the two variables X and Y is -1 , such a correlation is called perfect negative correlation. Let us consider an example to show the computation and interpretation of correlation coefficient.

Example 11.1: Consider the data on the quantity demanded and the price of a commodity over a ten-year period as given in the following table:

Year	Demand	Price
1996	100	5
1997	75	7
1998	80	6
1999	70	6
2000	50	8
2001	65	7
2002	90	5
2003	100	4
2004	110	3
2005	60	9

Estimate the correlation coefficient between the quantity demanded and price and interpret the same.

Solution:

This problem will be attempted first by showing all the detailed computations using the following formula.

$$r_{xy} = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X} \bar{Y}}{\sqrt{\sum_{i=1}^n X_i^2 - n\bar{X}^2} \sqrt{\sum_{i=1}^n Y_i^2 - n\bar{Y}^2}}$$

The required computations are shown in the following table:

Year	Demand (Y)	Price (X)	XY	X ²	Y ²
1996	100	5	500	25	10000
1997	75	7	525	49	5625
1998	80	6	480	36	6400
1999	70	6	420	36	4900
2000	50	8	400	64	2500
2001	65	7	455	49	4225
2002	90	5	450	25	8100
2003	100	4	400	16	10000
2004	110	3	330	9	12100
2005	60	9	540	81	3600
TOTAL	800	60	4500	390	67450

NOTES

NOTES

$$\begin{aligned}\Sigma XY &= 4500 & \Sigma X^2 &= 390 \\ \Sigma Y^2 &= 67,450 & \Sigma Y &= 800 \\ \Sigma X &= 60 & n &= 10\end{aligned}$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{800}{10} = 80 \quad \bar{X} = \frac{\Sigma X}{n} = \frac{60}{10} = 6$$

Substituting these values in the formula for the correlation coefficient, we get:

$$\begin{aligned}r_{xy} &= \frac{4500 - 10 \times 6 \times 80}{\sqrt{390 - 10 \times 6 \times 6} \sqrt{67450 - 64000}} \\ &= \frac{4500 - 4800}{\sqrt{390 - 360} \sqrt{67450 - 64000}} \\ &= \frac{-300}{\sqrt{30} \sqrt{3450}} = \frac{-300}{5.477 \times 58.737} \\ &= \frac{-300}{321.701} = -0.9325\end{aligned}$$

The value of the correlation coefficient between the quantity demanded and price is -0.9325 , which is negative and very high. This shows that the quantity demanded and price move in the opposite directions.

11.5 CORRELATION ANALYSIS

Correlation and regression analysis are generally performed together. Correlation measures the degree of the association between two or more set of variables. Regression, on the other hand, is used to explain the variations in one variable—usually called the dependent variable—by a set of independent variables. It identifies the nature of the relationship. The number of independent variables in regression analysis could be one or more. In case of one independent variable, we classify it as a simple regression, whereas in case of more than one independent variable, it is called a multiple regression analysis.

Correlation

Correlation measures the degree of association between two or more variables. When we are dealing with two variables, we are talking in terms of simple correlation and when more than two variables are involved, the subject matter of interest is called multiple correlation. In this section, we will start the discussion of simple correlation and extend the analysis to multiple correlation. There are three types of correlation:

1. Positive correlation: When two variables X and Y move in the same direction, the correlation between the two is positive. If one variable increases, the

other variable also increases and if one variable decreases, the other variable also decreases.

The examples of positive correlation are a particular quantity supplied of a commodity and the price of the commodity, the sales revenue and the advertising expenditure, consumption expenditure and the disposable income. The scatter of the points of the variables X and Y is clustered around a positively sloped line/curve in such a case as shown in Figure 11.1. In the figure, we note that the two variables X and Y move in the same direction.

NOTES

2. Negative correlation: When two variables X and Y move in the opposite direction, the correlation is negative. If one variable increases, the other decreases and vice versa. The examples of negative correlation are usually the quantity demanded and the price of the commodity. The scatter of the points on the variables X and Y is clustered around a negatively sloped straight line/curve in such a situation as shown in Figure 11.2. In the figure, we find that the variables X and Y are moving in the opposite direction.

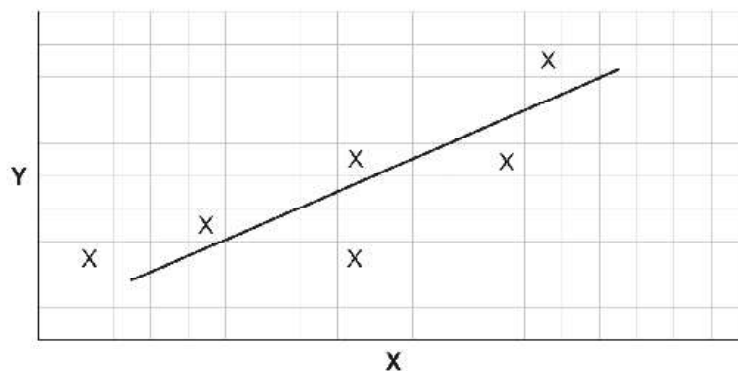


Figure 11.1 Positive correlation

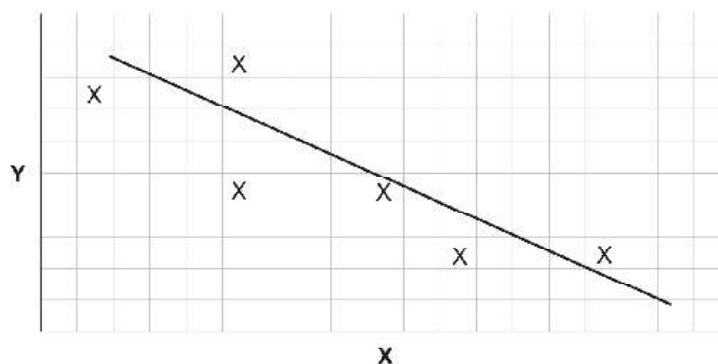


Fig. 11.2 Negative correlation

3. Zero correlation: The correlation between two variables X and Y is zero when the variables move in no connection with each other. If the variable X increases, Y may increase or decrease in some situation. The scatter of the points of the variables

X and Y in case of zero correlation is given in Figure 11.3. Zero correlation does not mean that the variables are not related. We are, here, dealing with a linear correlation and there could be a non-linear relation between them.

NOTES

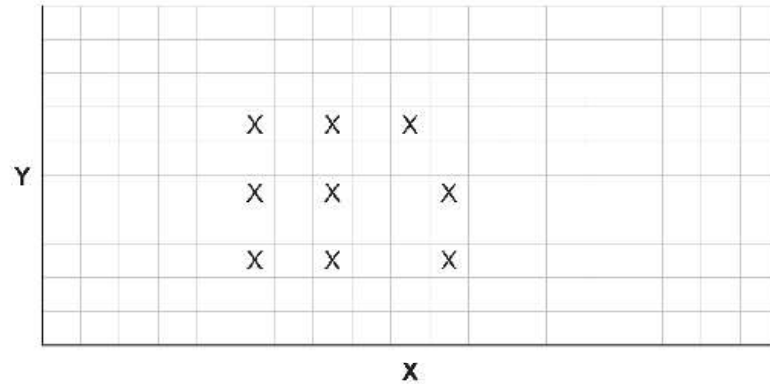


Fig. 11.3 Zero correlation

Quantitative Estimate of a Linear Correlation

A quantitative estimate of a linear correlation between two variables X and Y is given by Karl Pearson as:

$$r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

which may be rewritten as:

$$r_{xy} = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sqrt{\sum_{i=1}^n X_i^2 - n \bar{X}^2} \sqrt{\sum_{i=1}^n Y_i^2 - n \bar{Y}^2}}$$

Where, r_{xy} = Correlation coefficient between X and Y

$\bar{X}$ = Mean of the variable X

$\bar{Y}$ = Mean of the variable Y

n = Size of the sample

It may be noted that the above-mentioned formulae are for the linear correlation coefficient. The linear correlation coefficient takes a value between -1 and $+1$ (both values inclusive). If the value of the correlation coefficient is equal to 1 , the two variables are perfectly positively correlated and the scatter of the points of the variables X and Y will lie on a positively sloped straight line. Similarly, if the correlation coefficient between the two variables X and Y is -1 , the scatter of the

points of these variables will lie on a negatively sloped straight line and such a correlation will be called a perfectly negative correlation. It may be noted that the closer the scatter of points to the line, higher is the degree of correlation between the variables.

Testing the Significance of the Correlation Coefficient

The statistical test for the significance of a correlation coefficient is conducted using a t-statistic. The hypothesis to be tested is mentioned below:

$$H_0 : \rho = 0 \quad H_1 : \rho \neq 0$$

Test statistic is given by,

$$t_{n-2} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Where, r = Population correlation coefficient between the variables X and Y

r = Sample correlation coefficient between the variables X and Y

$n - 2$ = The degrees of freedom

Given the value of r and n , the value of the test statistic t could be computed. Now for a given level of significance, if computed $|t|$ is greater than tabulated $|t|$ with $n - 2$ degrees of freedom, the null hypothesis of no correlation between X and Y is rejected.

11.6 ANALYSIS OF VARIANCE

We have discussed the test of hypothesis concerning the equality of two population means using both the Z and t tests in Unit 8. However, if there are more than two populations, the test for the equality of means could be carried out by considering two populations at a time. This would be a very cumbersome procedure. One easy way out could be to use the Analysis of Variance (ANOVA) technique. The technique helps in performing this test in one go and, therefore, is considered to be important technique of analysis for the researcher. Through this technique it is possible to draw inferences whether the samples have been drawn from populations having the same mean.

The technique has found applications in the fields of economics, psychology, sociology, business and industry. It becomes handy in situations where we want to compare the means of more than two populations. Some examples could be to compare:

- the mean cholesterol content of various diet foods
- the average mileage of, say, five automobiles

NOTES

NOTES

- the average telephone bill of households belonging to four different income groups and so on.

R A Fisher developed the theory concerning ANOVA. The basic principle underlying the technique is that the total variation in the dependent variable is broken into two parts—one which can be attributed to some specific causes and the other that may be attributed to chance. The one which is attributed to the specific causes is called the variation between samples and the one which is attributed to chance is termed as the variation within samples. Therefore, in ANOVA, the total variance may be decomposed into various components corresponding to the sources of the variation.

In ANOVA, the dependent variable in question is metric (interval or ratio scale), whereas the independent variables are categorical (nominal scale). If there is one independent variable (one factor) divided into various categories, we have one-way or one-factor analysis of variance. In the two-way or two-factor analysis of variance, two factors, each divided into the various categories are involved.

In ANOVA, it is assumed that each of the samples is drawn from a normal population and each of these populations has an equal variance. Another assumption that is made is that all the factors except the one being tested are controlled (kept constant). Basically, two estimates of the population variances are made. One estimate is based upon between the samples and the other one is based upon within the samples. The two estimates of variances can be compared for their equality using F statistic.

Completely Randomized Design in a One-way ANOVA

Completely randomized design involves the testing of the equality of means of two or more groups. In this design, there is one dependent variable and one independent variable. The dependent variable is metric (interval/ratio scale) whereas the independent variable is categorical (nominal scale). A sample is drawn at random from each category of the independent variable. The size of the sample from each category could be equal or different. Let us consider a few examples to illustrate a one-way analysis of variance.

Example 11.2: Suppose we want to compare the cholesterol contents of the four competing diet foods on the basis of the following data (in milligrams per package) which were obtained for three randomly taken 6-ounce packages of each of the diet foods:

Diet Food A	3.6	4.1	4.0
Diet Food B	3.1	3.2	3.9
Diet Food C	3.2	3.5	3.5
Diet Food D	3.5	3.8	3.8

We want to test whether the difference among the sample means can be attributed to chance at the 5 per cent level of significance.

Solution: As explained earlier, the total variation in the data set can be expressed as a sum of the variations that can be attributed to specific sources (in this example, the various diet foods) plus the one which is attributed due to chance. The total variation in the data set is called the total sum of squares (TSS) and is computed as:

$$\text{TSS} = \sum_{i=1}^k \sum_{j=1}^n x_{ij}^2 - \frac{1}{kn} \cdot T_{..}^2$$

Where, $(i = 1, \dots, k \text{ and } j = 1, 2, \dots, n)$

x_{ij} = the j^{th} observation of the i^{th} sample (diet food)

$T_{..}$ = Grand total of all the data

$k = 4$ (Number of diet foods)

$n = 3$ (number of observations in each sample)

The term $\frac{1}{kn} \cdot T_{..}^2$ is referred to as the correction factor. The variation between the sample means which is attributed to specific sources or causes is referred to as the treatment sum of squares (TrSS). This is computed using the following formula:

$$\text{TrSS} = \frac{1}{n} \sum_{i=1}^k T_{i.}^2 - \frac{1}{kn} \cdot T_{..}^2$$

Where, $T_{i.}$ = Total of observations for the i^{th} treatment.

The variation within the sample, which is attributed to chance, is referred to as the error sum of squares (SSE). This could be computed by subtracting the treatment sum of squares from the total sum of squares. This is shown as:

$$\text{SSE} = \text{TSS} - \text{TrSS}$$

$$= \left[\sum_{i=1}^k \sum_{j=1}^n x_{ij}^2 - \frac{1}{kn} \cdot T_{..}^2 \right] - \left[\frac{1}{n} \sum_{i=1}^k T_{i.}^2 - \frac{1}{kn} \cdot T_{..}^2 \right]$$

In order to test the null hypothesis,

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D$$

against the alternative hypothesis

$$H_1: \text{At least two means are not equal}$$

(Treatment means are not equal)

We test the equality of TrSS with SSE. The necessary workings required for this are presented in Table 11.10, which is called one-way analysis of the variance table. The first column of the table indicates the sources of variation. The second column lists the degrees of freedom. There are k treatments;

NOTES

NOTES

therefore the corresponding degrees of freedom are $k - 1$. Similarly, the total number of observations in the data set is kn and therefore, the corresponding degrees of freedom are $kn - 1$. The degrees of freedom for errors are obtained by subtracting from the total degrees of freedom, the degrees of freedom corresponding to the treatment, i.e., $(kn - 1) - (k - 1) = k(n - 1)$. The third column lists the sum of squares due to the various sources of variation. The fourth column lists the mean square due to treatment $MSTr = (TrSS/k - 1)$ and the mean square due to error $MSE = (SSE/k(n - 1))$ obtaining by dividing the corresponding sum of squares by their degrees of freedom. The last column indicates the F statistic given as the ratio of the two mean squares with $k - 1$ degrees of freedom for the numerator and $k(n - 1)$ degrees of freedom for the denominator. For a given level of significance, the computed F statistic is compared with the table value of F with $k - 1$ degrees of freedom in the numerator and $k(n - 1)$ degrees of the freedom for the denominator. If the computed F value is greater than the tabulated F value, the null hypothesis is rejected.

Table 11.10 One-way ANOVA

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	$F_{k-1, k(n-1)}$
Treatments (Diet food)	$k - 1$	TrSS	$MSTr = \frac{TrSS}{k - 1}$	$\frac{MSTr}{MSE}$
Error	$k(n - 1)$	SSE	$MSE = \frac{SSE}{k(n - 1)}$	
Total	$kn - 1$	TSS		

The required computations in case of Example 11.2 are given below:

$$k = 4, n = 3$$

$T_{..}$	=	$3.6 + 4.1 + 4.0 + 3.1 + 3.2 + 3.9 + 3.2 + 3.5 + 3.5 + 3.5 + 3.8 + 3.8$	=	43.2
$T_{1.}$	=	$3.6 + 4.1 + 4.0$	=	11.7
$T_{2.}$	=	$3.1 + 3.2 + 3.9$	=	10.2
$T_{3.}$	=	$3.2 + 3.5 + 3.5$	=	10.2
$T_{4.}$	=	$3.5 + 3.8 + 3.8$	=	11.1
$\sum_{i=1}^4 \sum_{j=1}^3 x_{ij}^2$	=	$(3.6)^2 + (4.1)^2 + (4.0)^2 + (3.1)^2 + (3.2)^2 + (3.9)^2 + (3.2)^2 + (3.5)^2 + (3.5)^2 + (3.5)^2 + (3.8)^2 + (3.8)^2$	=	156.70

$$TSS = \sum_{i=1}^4 \sum_{j=1}^3 x_{ij}^2 - \frac{1}{kn} \cdot T_{..}^2 = 156.70 - \frac{1}{12} (43.2)^2 = 1.18$$

$$\begin{aligned}
 \text{TrSS} &= \frac{1}{n} \sum_{i=1}^4 T_{i\cdot}^2 - \frac{1}{kn} \cdot T_{..}^2 \\
 &= \frac{1}{3} [11.7^2 + 10.2^2 + 10.2^2 + 11.1^2] - \frac{1}{12} (43.2)^2 = 0.54 \\
 \text{SSE} &= \text{TSS} - \text{TrSS} \\
 &= 1.18 - 0.54 = 0.64
 \end{aligned}$$

The above results corresponding to Example 11.2 could be set up in the ANOVA Table 11.11.

Table 11.11 ANOVA Table for Example 11.2

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	$F_{\frac{3}{8}}$
Treatments (Diet Food)	3	0.54	0.18	2.25
Error	8	0.64	0.08	
Total	11	1.18		

Assuming the level of significance to be 5 per cent, the table value of F with 3 degrees of freedom in the numerator and 8 degrees of freedom in the denominator equals 4.07 (See Table 11.12). Since the computed F is less than the tabulated F , there is not enough evidence to reject the null hypothesis. Therefore, the difference in the cholesterol contents in the four diet foods could be attributed to chance.

As mentioned earlier, the size of the sample from each category (treatment) need not be same. If there are n_i observations corresponding to i^{th} treatment, the computing formula for the sum of squares would look like:

$$\text{TSS} = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}^2 - \frac{1}{N} \cdot T_{..}^2$$

$$\text{TrSS} = \sum_{i=1}^k \frac{T_{i\cdot}^2}{n_i} - \frac{1}{N} T_{..}^2$$

$$\text{SSE} = \text{TSS} - \text{TrSS}$$

Where, $N = n_1 + n_2 + \dots + n_k$

The total number of degrees of freedom in the case is $N - 1$, and the degrees of freedom are $k - 1$ for the treatments and $N - k$ for the error. Let us consider an example.

NOTES

Table 11.12(a) Significance points of the variance-ratio 'F' 5 per cent points of F**NOTES**

$\begin{matrix} v_1 \rightarrow \\ v_2 \downarrow \end{matrix}$	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	3.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.89
120	2.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

 v_1 = Degrees of freedom for greater variance. v_2 = Degrees of freedom for smaller variance.

Table 11.12(b) Significance points of the variance-ratio 'F' 1 per cent points of F

Data Analysis II

$v_1 \rightarrow$ $v_2 \downarrow$	1	2	3	4	5	6	8	12	24	∞
1	4052	5000	5403	5625	5764	5859	5982	6106	6235	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.37	99.42	99.46	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.13
4	21.20	18.20	16.69	15.88	15.52	15.21	14.80	14.37	13.93	13.45
5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	9.89	9.47	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.12	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.30	3.96	3.59	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.14	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.46	3.08	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.59
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.20	2.96	2.58	2.10
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.45	2.13
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
∞	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

v_1 = Degrees of freedom for greater variance.

v_2 = Degrees of freedom for smaller variance.

NOTES

NOTES

Example 11.3: The following are the number of kilometres/litre which a test driver with three different types of cars has obtained randomly on different occasions.

Car 1	15	14.5	14.8	14.9		
Car 2	13	12.5	13.6	13.8	14	
Car 3	12.8	13.2	12.7	12.6	12.9	13

Using a 5 per cent level of significance, perform a one-way ANOVA to examine the hypothesis that the difference in the average mileage in the three types of cars can be attributed to chance.

Solution:

$H_0 : \mu_1 = \mu_2 = \mu_3$ (Average mileage in the three types of cars is the same.)

H_1 : At least two types of cars do not have the same mileage.

$K = 3, n_1 = 4, n_2 = 5, n_3 = 6$

$N = n_1 + n_2 + n_3 = 4 + 5 + 6 = 15$

$T_{..}$	=	$15 + 14.5 + 14.8 + 14.9 + 13 + 12.5 + 13.6 + 13.8 + 14 + 12.8 + 13.2 + 12.7 + 12.6 + 12.9 + 13$	=	203.3
$T_{1.}$	=	$15 + 14.5 + 14.8 + 14.9$	=	59.2
$T_{2.}$	=	$13 + 12.5 + 13.6 + 13.8 + 14$	=	66.9
$T_{3.}$	=	$12.8 + 13.2 + 12.7 + 12.6 + 12.9 + 13$	=	77.2
$\sum_{i=1}^3 \sum_{j=1}^{n_i} x_{ij}^2$	=	$(15)^2 + (14.5)^2 + (14.8)^2 + (14.9)^2 + (13)^2 + (12.5)^2 + (13.6)^2 + (13.8)^2 + (14)^2 + (12.8)^2 + (13.2)^2 + (12.7)^2 + (12.6)^2 + (12.9)^2 + (13)^2$	=	2766.49

$$TSS = \sum_{i=1}^3 \sum_{j=1}^{n_i} x_{ij}^2 - \frac{1}{N} \cdot T_{..}^2$$

$$= 2766.49 - \frac{1}{15} (203.3)^2$$

$$= 2766.49 - 2755.393$$

$$= 11.097$$

$$TrSS = \sum_{i=1}^3 \frac{T_{i.}^2}{n_i} - \frac{1}{N} T_{..}^2$$

$$= \left[\frac{59.2^2}{4} + \frac{66.9^2}{5} + \frac{77.2^2}{6} \right] - \frac{1}{15} (203.3)^2$$

$$= 2764.5886 - 2755.3926$$

$$= 9.196$$

$$\begin{aligned}
 \text{SSE} &= \text{TSS} - \text{TrSS} \\
 &= 11.097 - 9.196 \\
 &= 1.901
 \end{aligned}$$

The ANOVA table in the case of Example 11.3 can be set up as shown in Table 11.13.

Table 11.13 One-way ANOVA for Example 11.3

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F_{12}^2
Treatments (Between groups)	2	9.196	4.598	29.02
Error (within groups)	12	1.901	0.158	
Total	14	11.097		

The computed F statistics equals 29.02. The table value of F with 2 degrees of freedom in the numerator and 12 degrees of freedom in the denominator at a 5 per cent level of significance is given by 3.89. As the computed F statistic is greater than the table F value, the null hypothesis is rejected. Therefore, the average mileage in these types of cars is statistically different.

Randomized Block Design in Two-way ANOVA

In Example 11.2, it could not be shown that there really is a significant difference in the average cholesterol content of the four diet foods. The results were not statistically different because there was a considerable difference in the values within each of the samples resulting in a large experimental error. However, if we have additional information that each of the value was randomly measured in the three different laboratories in such a way that the first value of each sample came from laboratory 1, the second value from laboratory 2, and the third value from laboratory 3 (the random assignment of test units to labs). In such a case, a two-way analysis of variance is suggested. We had earlier partitioned the total sum of squares into two components—one which is due to the differences between the sample (treatment sum of squares) and the other one due to the differences within the samples (error sum of squares). Now, this error sum of square includes the sum of squares due to laboratories (called blocks) as an extraneous factor. In two-way analysis of variance, we remove the effect of the extraneous factors (laboratories or blocks) from the error sum of squares. Therefore, the total sum of square is partitioned into three components—one due to treatment, second due to block and the third one due to chance (called the error sum of squares). It may be noted that the total sum of squares (TSS) and the treatment sum of squares (TrSS) would remain the same as computed earlier in Example 11.2. In addition, we will have another component called block sum of squares (SSB) which is due to different laboratories and is computed as:

$$\text{SSB} = \frac{1}{k} \cdot \sum_{j=1}^n T_{\cdot j}^2 - \frac{1}{kn} \cdot T_{..}^2$$

NOTES

Where, $T_{.j}$ = Total of the values in the j^{th} block.

The error sum of squares would be computed as:

$$SSE = TSS - \text{TrSS} - \text{SSB}$$

NOTES

There will be two hypotheses to be tested:

I (Diet Food)

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D$$

H_1 : At least the two means are not same.

II (Blocks or Labs)

$$H_0 : v_1 = v_2 = v_3$$

(Average cholesterol content in the three labs is same.)

H_1 : At least two means are not same.

Now, we would need to test the equality of TrSS with SSE and SSB with SSE. The necessary working required for this are presented in Table 11.14 called two-way analysis of variance table.

Table 11.14 Two-way ANOVA

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatments	$k - 1$	TrSS	$MSTr = \frac{\text{TrSS}}{k - 1}$	$F_{(k-1)(n-1)}^{k-1} = \frac{MSTr}{MSE}$
Blocks	$n - 1$	SSB	$MSB = \frac{\text{SSB}}{n - 1}$	$F_{(k-1)(n-1)}^{n-1} = \frac{MSB}{MSE}$
Error	$(k - 1)(n - 1)$	SSE	$MSE = \frac{\text{SSE}}{(k - 1)(n - 1)}$	
Total	$kn - 1$	TSS		

Example 11.2 can be rewritten as Example 11.4.

Example 11.4: Suppose in Example 11.2, the measurement of the cholesterol content was performed in three different laboratories. The first value of each sample came from one laboratory, the second value came from another laboratory, and the third value came from a third laboratory. The data is presented below:

Diet Food	Laboratory		
	One	Two	Three
Diet Food A	3.6	4.1	4.0
Diet Food B	3.1	3.2	3.9
Diet Food C	3.2	3.5	3.5
Diet Food D	3.5	3.8	3.8

Perform a two-way ANOVA using a 0.05 level of significance.

Solution: There will be two hypotheses to be tested in this case; one corresponding to the treatment (diet food) and the other corresponding to laboratories (blocks). These are listed below:

I (Diet Food)

H_0 : $\mu_A = \mu_B = \mu_C = \mu_D$ (Average cholesterol content of the four diet foods is same.)

H_1 : At least two means are not same.

II (Blocks or labs)

H_0 : $v_1 = v_2 = v_3$ (Average cholesterol content in the three labs is same.)

H_1 : At least two means are not same.

The TSS and TrSS here would be the same as computed in Example 11.2. As mentioned earlier, the block sum of square would be required in this problem using the formula:

$$SSB = \frac{1}{k} \cdot \sum_{j=1}^n T_{\bullet j}^2 - \frac{1}{kn} \cdot T_{\bullet\bullet}^2$$

Where, $T_{\bullet j}$ = Total of the values in the j^{th} block.

The error sum of squares would be obtained as

$$SSE = TSS - \text{TrSS} - SSB$$

The required computations for the two-way ANOVA are as under:

$$T_{\bullet 1} = 3.6 + 3.1 + 3.2 + 3.5 = 13.4$$

$$T_{\bullet 2} = 4.1 + 3.2 + 3.5 + 3.8 = 14.6$$

$$T_{\bullet 3} = 4.0 + 3.9 + 3.5 + 3.8 = 15.2$$

$$SSB = \frac{1}{k} \cdot \sum_{j=1}^n T_{\bullet j}^2 - \frac{1}{kn} \cdot T_{\bullet\bullet}^2$$

$$= \frac{1}{4} [13.4^2 + 14.6^2 + 15.2^2] - \frac{1}{12} (43.2)^2$$

$$= 155.94 - 155.52$$

$$= 0.42$$

We have already computed in Example 12.1, the values of TSS and TrSS as under:

$$TSS = 1.18, \text{TrSS} = 0.54$$

$$\text{Therefore, } SSE = TSS - \text{TrSS} - SSB$$

$$= 1.18 - 0.54 - 0.42$$

$$= 0.22$$

NOTES

NOTES

We note that the SSE in Example 11.2 was 0.64, whereas here it is 0.22. This is because the earlier SSE has been partitioned into two components, namely, the block sum of squares (SSB) having a value of 0.42 resulting in 0.22 as the new error sum of squares (SSE). The required results for the testing of the two hypotheses are presented in the ANOVA Table 11.15.

Table 11.15 Two-way ANOVA Table for Example 11.4

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	<i>F</i>
Treatments (Diet Food)	3	0.54	0.18	$F_6^3 = \frac{0.18}{0.0367} = 4.90$
Block (Laboratories)	2	0.42	0.21	$F_6^2 = \frac{0.21}{0.0367} = 5.72$
Error (Chance)	6	0.22	0.0367	
Total	11	1.18		

The table value of F_6^3 and F_6^2 at a 5 per cent level of significance is given by 4.76 and 5.14 respectively. The corresponding sample *F* values for both are 4.90 and 5.72. Since the computed *F* values are greater than the corresponding table values, the null hypothesis is rejected in both the cases. Therefore, it can be concluded that there is a difference in the average cholesterol content due to various diet foods and because of the laboratories where the measurements were taken.

Factorial Design

In factorial design, the dependent variable is the interval or the ratio scale and there are two or more independent variables which are nominal scale. In the factorial design, it is possible to examine the interaction between the variables. If there are two independent variables each having three cells, there would be a total of nine interactions. The details on this are already explained in Unit 3 (Research Design). The main advantage of factorial design over randomized block design is that it is possible to measure the main effects as well as the interaction effects of two or more independent variables at various levels. Further, the randomized block design has only two independent variables whereas, factorial design can take care of more than two independent variables. Let us consider an illustration to explain factorial design.

It is generally observed that there are differences in the pay packages offered to fresh MBA graduates. The variations could be either due to the type of business school where they have studied or it could be due to their area of specialization. The variation can also be due to an interaction between the business school and the area of specialization. For example, the specialization in finance at one business school might fetch a better package. All these presumptions could be tested with the help of the factorial design explained with the help of the following example.

Example 11.5: The following data refers to the salary package (in ₹lakhs) offered to MBA graduates with different specializations and having studied at four different

business schools. For the sake of simplification, only two students are taken for each interaction between the institute and field of specialization.

Specialization	Business School			
	I	II	III	IV
Marketing	6	4	8	6
	5	5	6	4
Finance	7	6	6	9
	6	7	7	8
Operations	8	5	10	9
	7	5	9	10

NOTES

Test the hypothesis (i) whether the difference between the pay packages offered can be attributed to chance (ii) average pay packages by all specializations are equal. (iii) the average pay package for 12 interactions are equal.

You may use a 5 per cent level of significance.

Solution: The following set of hypotheses is required to be tested.

Business schools:

H_0 : Average pay package for all the institutions are equal.

H_1 : Average pay package for all the institutions are not equal

Specialization:

H_0 : Average pay package for all the specializations are equal.

H_1 : Average pay package for all the specializations are not equal

Interaction:

H_0 : Average pay package for all 12 interactions are equal.

H_1 : Average pay package for all 12 interactions are not equal

Let us compute the following:

$$\begin{aligned} \text{Correction factor (CF)} &= \frac{(\text{Sum of all observations})^2}{\text{Total number of observations}} \\ &= \frac{(163)^2}{24} = \frac{26569}{24} = 1107.04 \end{aligned}$$

$$\begin{aligned} \text{Total sum of squares} &= (\text{Sum of squares of observations}) - \text{CF} \\ &= 6^2 + 4^2 + 8^2 + 6^2 + \dots + 7^2 + 5^2 + 9^2 + 10^2 - 1107.04 \\ &= 1179 - 1107.04 \\ &= 71.96 \end{aligned}$$

Sum of squares due to specialization (row)/SSR

$$\begin{aligned}
 &= \frac{44^2}{8} + \frac{56^2}{8} + \frac{63^2}{8} - CF \\
 &= 1130.13 - 1107.04 \\
 &= 23.08
 \end{aligned}$$

NOTES

Where,

$$\text{Sum total for Marketing} = 44$$

$$\text{Sum total for Finance} = 56$$

$$\text{Sum total for Operations} = 63$$

Sum of squares due to school (column)/SSC

$$\begin{aligned}
 &= \frac{39^2}{6} + \frac{32^2}{6} + \frac{46^2}{6} + \frac{46^2}{6} - CF \\
 &= 1129.5 - 1107.04 \\
 &= 22.46
 \end{aligned}$$

Where,

$$\text{Sum total for Business School I} = 39$$

$$\text{Sum total for Business School II} = 32$$

$$\text{Sum total for Business School III} = 46$$

$$\text{Sum total for Business School IV} = 46$$

$$\text{Sum of squares due to interactions (SSI)} = n \sum (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2$$

Where,

n = number of observations for each interaction

$\bar{x}_{i.}$ = Mean of observations of i^{th} row

$\bar{x}_{.j}$ = Mean of observation of j^{th} column

$\bar{x}_{..}$ = Grand mean of all the observations.

$\bar{x}_{ij}$ = Mean of observation of i^{th} row and j^{th} column

The above terms can be calculated by first calculating the means of all the interactions and also the means of the corresponding rows and columns. These are presented in the table below:

Specialization	Business School				$\bar{x}_{i.}$
	I	II	III	IV	
Marketing	5.5	4.5	7	5	5.5
Finance	6.5	6.5	6.5	8.5	7
Operations	7.5	5	9.5	9.5	7.88
$\bar{x}_{.j}$	6.5	5.33	7.67	7.67	$\bar{x}_{..} = 6.793$

Therefore,

$$\begin{aligned}
 \text{SSI} &= 2 \sum \sum (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2 \\
 &= 2[5.5 - 5.5 - 6.5 + 6.79]^2 + (4.5 - 5.5 - 5.33 + 6.79)^2 \\
 &\quad + (9.5 - 7.88 - 7.67 + 6.79)^2 \\
 &= 2 \times 8.96 = 17.92
 \end{aligned}$$

Sum of Squares due to error (SSE):

$$\begin{aligned}
 \text{SSE} &= \text{TSS} - \text{SSR} - \text{SSC} - \text{SSI} \\
 &= 71.96 - 23.08 - 22.46 - 17.92 \\
 &= 8.5
 \end{aligned}$$

Therefore, the ANOVA table for factorial design could be prepared as given in Table 11.16.

Table 11.16 Results of ANOVA Table for Factorial Design

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Sum of Squares	F
Row (Specialization)	23.08	2	11.54	16.26
Column (Business School)	22.46	3	7.49	10.55
Interaction	17.92	6	2.96	4.17
Error	8.50	12	0.71	
Total	71.96	23		

The table values of F_{12}^2 , F_{12}^3 and F_{12}^6 (at 5 per cent level of significance) are given as 3.885, 3.490 and 2.996 respectively. As the computed value for the hypothesis concerning specialization, business school and interaction are greater than the corresponding tabulated values; the three null hypotheses are rejected. This means that it can be concluded that the packages offered to the graduates vary due to their specialization, the type of business school in which they have studied and their interactions.

Check Your Progress

- What is frequency distribution?
- What is regression used for?
- What is the basic principle underlying the ANOVA technique?
- What is the main advantage of factorial design over randomized block design?

NOTES

11.7 ANALYSIS OF COVARIANCE

NOTES

Analysis of covariance, represents an extension of analysis of variance that tests the significance of difference between means of final experimental data by taking into account the correlation between the dependent variable and one or more co-variates or pertinent control variables, and by adjusting initial mean differences in the groups. It is useful to experimental psychologists, especially in the field of memory and learning, when for various reasons it is not possible or quite difficult to equate experimental groups on some pertinent variable at the start of the experiment.

In an experiment, suppose the dependent variable was the performance of two groups of students on a criterion test; one group was taught through programmed learning (experimental group) and the other through conventional method (control group). The two groups were framed by randomly assigning 70 subjects to each of the groups and the subjects were equated on the pre-test scores on the criterion test. Since intelligence was also considered to be a significant factor related to the scores on the criterion test, the scores on the intelligence test were also taken into account so as to provide a statistical control for testing the significance of the difference between the means of criterion test scores of the two groups at the end of the experiment.

Let scores of the subjects of the experimental and control groups on intelligence test be denoted by X_1 and X_2 and those on the criterion test (post test) as Y_1 and Y_2 respectively. The total number of subjects in experimental and control groups be denoted by N_1 and N_2 respectively. Moreover, the sum of the scores and the sum of squares of the scores on intelligence and criterion tests obtained by the subjects of the two groups be denoted by the following symbols:

ΣX_1 = sum of the scores on intelligence test obtained by the subjects of the experimental group.

ΣX_2 = sum of the scores on intelligence test obtained by the subject of the control group.

$$\Sigma X = \Sigma X_1 + \Sigma X_2$$

ΣX_1^2 = sum of the squares of the scores on intelligence test obtained by the subjects of the experimental group.

ΣX_2^2 = the sum of the squares of the scores on intelligence test obtained by the subject of the control group.

$$\Sigma X^2 = \Sigma X_1^2 + \Sigma X_2^2$$

ΣY_1 = sum of the scores on criterion test obtained by subjects of the experimental group.

ΣY_2 = sum of the scores on criterion test obtained by the subjects of the control group.

$$\Sigma Y = \Sigma Y_1 + \Sigma Y_2$$

ΣY_1^2 = sum of the squares of the scores on criterion test obtained by the subjects of the experimental group.

ΣY_2^2 = sum of the squares of the scores on criterion test obtained by the subjects of the control group.

$$\Sigma Y^2 = \Sigma Y_1^2 + \Sigma Y_2^2$$

$\Sigma X_1 Y_1$ = sum of the products of the scores on intelligence and criterion test obtained by the subjects of the experimental group.

$\Sigma X_2 Y_2$ = sum of the products of the scores on intelligence test and criterion test obtained by the subjects of control group.

$$\Sigma XY = \Sigma X_1 Y_1 + \Sigma X_2 Y_2$$

N = total number of subjects in the experimental and control groups

$$= N_1 + N_2$$

The outcomes of the experiment are given in Table 11.17

$$\begin{array}{lll} \Sigma X = \Sigma X_1 + \Sigma X_2 & \Sigma XY = \Sigma X_1 Y_1 + \Sigma X_2 Y_2 & \Sigma Y = \Sigma Y_1 + \Sigma Y_2 \\ = 3188 + 3283 & = 257968 + 256161 & = 5608 + 5421 \\ = 6471 & = 514129 & = 11029 \\ \Sigma X^2 = \Sigma X_1^2 + \Sigma X_2^2 & & \Sigma Y^2 = \Sigma Y_1^2 + \Sigma Y_2^2 \\ = 151488 + 161359 & & = 451038 + 422299 \\ = 312847 & & = 873337 \end{array}$$

Table 11.17. Data of the Experimental and Control Groups Obtained under Two-experimental Treatments

Experimental group			Control group		
Subject	Score on intelligence test (X_1)	Score on criterion test (Y_1)	Subject	Score on intelligence test (X_2)	Score on criterion test (Y_2)
1	50	84	1	37	70
2	35	80	2	47	81
3	28	76	3	46	70
4	43	80	4	58	84
5	47	83	5	39	80
6	41	81	6	42	76
7	29	70	7	52	81
8	46	82	8	30	65

NOTES

NOTES

9	29	68	9	44	70
10	28	76	10	32	72
11	43	81	11	50	82
12	46	82	12	49	81
13	35	78	13	52	70
14	38	81	14	57	80
15	39	81	15	40	69
16	49	80	16	24	80
17	41	81	17	43	80
18	46	81	18	38	80
19	43	80	19	46	83
20	45	82	20	40	78
21	59	85	21	54	81
22	46	82	22	52	75
23	43	81	23	48	80
24	38	70	24	53	68
25	49	83	25	48	81
26	59	80	26	46	70
27	53	84	27	44	80
28	52	84	28	40	72
29	48	83	29	60	85
30	56	84	30	47	81
31	42	81	31	48	80
32	33	68	32	70	62
33	54	84	33	52	80
34	58	85	34	27	60
35	63	85	35	49	81
36	58	84	36	49	77
37	57	84	37	50	72
38	60	86	38	51	81
39	23	70	39	42	72
40	53	80	40	24	62
41	63	86	41	68	84

42	48	83	42	46	80
43	44	81	43	47	78
44	40	80	44	60	84
45	46	70	45	61	75
46	55	84	46	59	81
47	26	75	47	52	70
48	48	83	48	53	80
49	51	81	49	27	70
50	42	80	50	43	82
51	24	63	51	50	77
52	53	74	52	38	78
53	44	80	53	39	80
54	51	83	54	61	86
55	52	84	55	60	85
56	47	82	56	49	84
57	48	82	57	51	80
58	50	83	58	63	81
59	47	82	59	19	70
60	56	85	60	51	79
61	39	70	61	47	82
62	46	82	62	38	80
63	62	86	63	40	77
64	37	80	64	49	80
65	46	81	65	28	78
66	53	84	66	50	80
67	39	80	67	56	84
68	46	81	68	59	86
69	49	83	69	52	79
70	31	70	70	50	79
$M_{X_1} =$		$M_{X_2} =$	$G_{M_X} =$	$M_{Y_1} =$	$M_{Y_2} =$
45.54		46.90	46.22	80.11	77.44

Analysis of co-variance has certain basic assumptions underlying it. According to Ray (1960, pp. 129-132), Lindquist (1953), and Edwards (1971, pp. 332-339) the assumptions are that:

Data Analysis II

NOTES

NOTES

1. The distribution of the adjusted scores within groups should be normal.
2. The methods groups should be selected at random from the same population.
3. The groups should be homogeneous in variability.
4. The effect to each treatment should be constant and additive on the responses of the subjects in the group on which the treatment is imposed.
5. There should be a real correlation between the dependent variable and the co-variate or pertinent control variable.
6. There should be homogeneity of regressions from group to group.

The basic assumptions underlying the technique of analysis of co-variance for the data Y_1 and Y_2 in Table 11.17 are tested as under:

- (i) **Assumption of normality.** In the light of the findings of Eden and Yates (Johnson, 1961), and Norton (Guilford, 1965, pp. 300-301) the assumption of normality may not be considered important for the data of the experiment.
- (ii) **Assumption of randomness.** The requirement of randomness was amply fulfilled for the data of the experiment. First 70 pairs of subjects were chosen on the basis of subject-to-subject matching on the pre-test criterion scores. Finally two groups, with 70 subjects in each group, were framed by randomly assigning one member of each pair to each group.
- (iii) **Assumption of homogeneity.** The assumption of homogeneity of sample variances is tested by applying Bartlett's test of homogeneity of variance.

1. Computation of $\Sigma Y_1 K^2$ and $\Sigma S_1 K^2$

$$\begin{aligned}\Sigma Y_1 K^2 &= \Sigma Y_1^2 - \frac{(\Sigma Y_1)^2}{N_1} \\ &= 451038 - \frac{(5608)^2}{70} \\ &= 1757.09\end{aligned}$$

$$\begin{aligned}\Sigma S_1 K^2 &= \frac{\Sigma Y_1 K^2}{N_1 - 1} \\ &= \frac{1757.09}{70 - 1} = 25.47\end{aligned}$$

2. Computation of $\Sigma Y_2 K^2$ and $\Sigma S_2 K^2$

$$\begin{aligned}\Sigma Y_2 K^2 &= \Sigma Y_2^2 - \frac{(\Sigma Y_2)^2}{N_2} \\ &= 422299 - \frac{(5421)^2}{70} \\ &= 2481.27\end{aligned}$$

$$\begin{aligned}\Sigma S_2 K^2 &= \frac{\Sigma Y_2 K^2}{N_2 - 1} \\ &= \frac{2481.27}{70 - 1} = 35.96\end{aligned}$$

3. *Computation of $K \log \frac{\Sigma SK^2}{2}$*

$$\begin{aligned}K \log \frac{\Sigma SK^2}{2} &= K \log \frac{\Sigma S_1 K^2 + \Sigma S_2 K^2}{2} \\ &= 2 \log \frac{25.47 + 35.96}{2} \\ &= 2 \log 30.715 \\ &= 2.9744\end{aligned}$$

4. *Computation of $\Sigma \log \Sigma SK^2$*

$$\begin{aligned}\Sigma \log \Sigma SK^2 &= \log \Sigma S_1 K^2 + \log \Sigma S_2 K^2 \\ &= \log 25.47 + \log 35.96 \\ &= 2.9618\end{aligned}$$

5. *Computation of difference D*

$$\begin{aligned}D &= K \log \frac{\Sigma SK^2}{2} - \Sigma \log \Sigma SK^2 \\ &= 2.9744 - 2.9618 = 0.0126\end{aligned}$$

6. *Computation of chi-square χ^2*

$$\begin{aligned}\chi^2 &= (2.3026) (N - 1)(D) \\ &= (2.3026) (70 - 1) (0.0126) \\ &= 2.002\end{aligned}$$

7. *Computation of correction C*

$$\begin{aligned}C &= 1 + \frac{K + 1}{3K \left(\frac{N_1 + N_2}{2} - 1 \right)} \quad (K \text{ is the number of treatment}) \\ &= 1 + \frac{2 + 1}{3 \times 2 \times \left(\frac{70 + 70}{2} - 1 \right)} \\ &= 1.007.\end{aligned}$$

8. *Computation of corrected chi-square*

$$\begin{aligned}\text{corrected } \chi^2 &= \frac{\chi^2}{\text{correction}(C)} \\ &= \frac{2.002}{1.007} = 1.988\end{aligned}$$

NOTES

$$9. df = K - 1 = 2 - 1 = 1$$

The value of c^2 needed for significance for 1 df at .05 level is 3.841. Since the obtained value 1.988 of c^2 is less than the needed value, the variance within the treatments may be treated as homogeneous.

NOTES

(iv) **Assumption of additivity.** The assumption that the effect of treatment is constant and additive on the responses of the subjects in the group on which the treatment was imposed is tested by the *Test of Additivity* in the manner described by Ray (1960, pp. 57-58). This test employs a variance ratio calculated in terms of F -value described as under:

1. *Sum of the squares for the experimental group*

$$\begin{aligned} &= \Sigma Y_1^2 - \frac{(\Sigma Y_1)^2}{N_1} \\ &= 451038 - \frac{(5608)^2}{70} = 1757.09 \end{aligned}$$

Variance for the experimental group

$$\begin{aligned} &= \frac{1757.09}{N_1 - 1} = \frac{1757.09}{70 - 1} \\ &= \frac{1757.09}{69} \end{aligned}$$

2. *Sum of the squares for the control group*

$$\begin{aligned} &= \Sigma Y_2^2 - \frac{(\Sigma Y_2)^2}{N_2} \\ &= 422299 - \frac{(5421)^2}{70} \\ &= 2481.27 \end{aligned}$$

Variance for the control group

$$= \frac{2481.27}{N_2 - 1} = \frac{2481.27}{70 - 1} = \frac{2481.27}{69}$$

Variance ratio

$$F = \frac{1757.09}{69} \bigg| \frac{2481.27}{69} = 0.71$$

Since the value of F needed for significance for 69/69 df at .05 level of confidence is greater than the obtained value 0.71 of F , the assumption of additivity is found to hold good for the data Y_1 and Y_2 in the Table 11.17.

(v) **Assumption of correlation.** The assumption that there should be a real correlation between the dependent variable and the pertinent control variable is tested by using the *Test of Significance of Regression* in the

manner described by Ray (1960, p. 129). This test also employs the computation of variance ratio in terms of F -value. In the present problem, the scores on the intelligence test are the pertinent control variable and those on the criterion test (post test) as the dependent variable.

The details of computational steps are described as under:

1. *The predicted sum of squares based on the within sample sums of squares and products*

$$P_w = \frac{\left[\Sigma X_1 Y_1 - \frac{(\Sigma X_1)(\Sigma Y_1)}{N_1} + \Sigma X_2 Y_2 - \frac{(\Sigma X_2)(\Sigma Y_2)}{N_2} \right]^2}{\left[\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{N_1} + \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{N_2} \right]}$$

$$= \frac{(2563.66 + 1916.10)^2}{(6297.37 + 7386.30)}$$

$$= \frac{(4479.76)^2}{13683.68} = 1466.58$$

2. *The predicted variance*

$$V_{pw} = \frac{P_w}{C-1} \quad (C \text{ is the number of groups})$$

$$= \frac{1466.58}{2-1} = 1466.58$$

3. *Within sample sum of squares of errors of estimate*

$$E_w = \left[\Sigma Y_1^2 - \frac{(\Sigma Y_1)^2}{N_1} + \Sigma Y_2^2 - \frac{(\Sigma Y_2)^2}{N_2} \right]$$

$$- \frac{\left[\Sigma X_1 Y_1 - \frac{(\Sigma X_1)(\Sigma Y_1)}{N_1} + \Sigma X_2 Y_2 - \frac{(\Sigma X_2)(\Sigma Y_2)}{N_2} \right]^2}{\left[\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{N_1} + \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{N_2} \right]}$$

$$= 1757.09 + 2481.27 - \frac{(2563.66 + 1916.10)^2}{(6297.37 + 7386.30)}$$

$$= 4238.36 - \frac{(4479.76)^2}{13683.67}$$

$$= 2771.78$$

4. *The error variance*

$$V_{ew} = \frac{E_w}{(N-C-1)} \quad (N = N_1 + N_2)$$

NOTES

$$= \frac{2771.78}{(140 - 2 - 1)} = 20.23$$

5. The variance ratio

$$F = \frac{V_{pw}}{V_{ew}}$$

$$= \frac{1466.58}{20.23} = 72.495$$

Since the obtained value 72.495 of F exceeds the value of F given in the Table D of the Appendix for 1/137 df at .05 level of confidence, the assumption of real correlation between the dependent variable and pertinent control variable is justified and that the use of intelligence scores as a pertinent control variable was worthwhile in partially eliminating error from the data.

(vi) **Assumption of homogeneity of regressions.** The assumption that there should be homogeneity of regressions from group-to-group is tested by computing the F-ratio for the *test of homogeneity of regressions* in the following manner as given by Ray (1960, pp. 130-131):

1. The sums of the squares of errors of prediction

$$E_i = \left[\Sigma Y_1^2 - \frac{(\Sigma Y_1)^2}{N_1} \right] - \frac{\left[\Sigma X_1 Y_1 - \frac{(\Sigma X_1)(\Sigma Y_1)}{N_1} \right]^2}{\left[\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{N_1} \right]}$$

$$+ \left[\Sigma Y_2^2 - \frac{(\Sigma Y_2)^2}{N_2} \right] - \frac{\left[\Sigma X_2 Y_2 - \frac{(\Sigma X_2)(\Sigma Y_2)}{N_2} \right]^2}{\left[\Sigma X_2^2 - \frac{(\Sigma X_2)^2}{N_2} \right]}$$

$$= 1757.09 - \frac{(2563.66)^2}{6297.37} + 2481.27 - \frac{(1916.10)^2}{7386.30}$$

$$= 713.42 + 1984.21 = 2697.63$$

2. *Within sample sum of squares of errors of estimate* (computed earlier in case of assumption of correlation by using formula.

$$E_w = 2771.78$$

3. The measure of homogeneity of regressions

$$= E_w - E_i$$

$$= 2771.78 - 2697.63$$

NOTES

$$= 74.15$$

4. The variance ratio

$$F = \frac{E_w - E_i}{C - 1} \bigg| \frac{E_i}{N - 2C} \quad (C \text{ is number of groups})$$

$$= \frac{74.15}{2 - 1} \bigg| \frac{2697.63}{140 - 2 \times 2}$$

$$= 3.74$$

Since the value of F needed for significance for 1/136 df at .05 level of confidence is greater than the obtained value 3.74 of F , the assumption that two sample regressions are homogeneous is taken to be justified.

NOTES

Check Your Progress

7. What does analysis of co-variance represent?
8. How is the assumption of homogeneity of sample variances tested?

11.8 MULTIVARIATE ANALYSIS

Let us now discuss multivariate analysis.

Factor Analysis

Factor analysis is a multivariate statistical technique in which there is no distinction between dependent and independent variables. In factor analysis, all variables under investigation are analyzed together to extract the underlined factors. Factor analysis is a data reduction method. It is a very useful method to reduce a large number of variables resulting in data complexity to a few manageable factors. These factors explain most part of the variations of the original set of data. A market researcher might have collected data on say, more than 50 attributes (or items) of a product which may become very difficult to analyze. Factor analysis could help to reduce the data on 50 odd attributes to a few manageable factors. It helps in identifying the underlying structure of the data.

A *factor* is a linear combination of variables. It is a construct that is not directly observable but that needs to be inferred from the input variables. The factors are statistically independent. We will show you their application in a regression analysis as the factor scores, when used as independent variables in regression analysis, help to solve the problem of multicollinearity. (The problem of multicollinearity in a regression model arises when the independent variables are so highly correlated that it becomes difficult to separate out the influence of each of the independent variables on the dependent variable.) The factor scores could also be used in other multivariate techniques.

Uses of Factor Analysis

The technique of factor analysis has multiple uses as discussed in the following situations:

NOTES

Scale construction: Factor analysis could be used to develop concise multiple item scales for measuring various constructs. We have already discussed in the chapter Attitude Measurement and Scaling the process of developing a multiple item scale that typically starts generating a large set of items (statements) relating to the attitude being measured. This is done as part of exploratory research. Factor analysis can reduce the set of statements to a concise instrument and at the same time, ensure that the retained statements adequately represent the critical aspects of the constructs being measured. Suppose we want to prepare a multiple item scale for measuring the job satisfaction of skilled workers in an organization. As the first step, we would generate a large number of statements, numbering say 100 or so as part of exploratory research. These statements could be subjected to factor analysis and let us assume that we get three factors out of it. Now, if we want to construct a 15-item scale to measure job satisfaction, what could be done is to separate five items in each of the factors having the highest factor loading. The concept of factor loading will be discussed later in the book. This way, a 15-item scale to measure job satisfaction could be developed.

Establish antecedents: This method reduces multiple input variables into grouped factors. Thus, the independent variables can be grouped into broad factors. For example, all the variables that measure the safety clauses in a mutual fund could be reduced to a factor called safety clause. Thus, the company could know about the broad benefit that an investor seeks in a fund.

Psychographic profiling: Different independent variables are grouped to measure independent factors. These are then used for identifying personality types. One of the most well-known inventories based on this technique is called the 16 PF inventory.

Segmentation analysis: Factor analysis could also be used for segmentation. For example, there could be different sets of two-wheelers-customers owning two wheelers because of different importance they give to factors like prestige, economy consideration and functional features.

Marketing studies: The technique has extensive use in the field of marketing and can be successfully used for new product development; product acceptance research, developing of advertising copy, pricing studies and for branding studies. For example we can use it to:

- identify the attributes of brands that influence consumers' choice;
- get an insight into the media habits of various consumers;
- identify the characteristics of price-sensitive customers.

Conditions for a Factor Analysis Exercise

Factor analysis requires some specific conditions that must be ensured before executing the technique. These are mentioned in detail in this section.

- Factor analysis exercise requires metric data. This means the data should be either interval or ratio scale in nature. The variables for factor analysis are identified through exploratory research which may be conducted by reviewing the literature on the subject, researches carried out already in this area, by informal interviews of knowledgeable persons, qualitative analysis like focus group discussions held with a small sample of the respondent population, analysis of case studies and judgement of the researcher. Generally in a survey research, a five or seven-point Likert scale or any other interval scales may be used.
- As the responses to different statements are obtained through different scales, all the responses need to be standardized. The standardization helps in comparison of different responses from such scales. The standardization is carried out using the following formulae:

Standardized score of i th respondent on a statement =

$$\frac{\text{Actual score of } i^{\text{th}} \text{ respondent on statement} - \text{Mean of all respondents on the statement}}{\text{Standard deviation of all respondents on the statement}}$$

- The size of the sample respondents should be at least four to five times more than the number of variables (number of statements).
- The basic principle behind the application of factor analysis is that the initial set of variables should be highly correlated. If the correlation coefficients between all the variables are small, factor analysis may not be an appropriate technique. A correlation matrix of the variables could be computed and tested for its statistical significance. The hypothesis to be tested may be written as:
 H_0 : Correlation matrix is insignificant, i.e., correlation matrix is an identity matrix where diagonal elements are one and off diagonal elements are zero.

H_1 : Correlation matrix is significant.

The test is carried out by using a Bartlett test of sphericity, which takes the determinant of the correlation matrix into consideration. The test converts it into a chi-square statistics with degrees of freedom equal to $[(k(k-1))/2]$, where k is the number of variables on which factor analysis is applied. The significance of the correlation matrix ensures that a factor analysis exercise could be carried out.

- Another condition which needs to be fulfilled before a factor analysis could be carried out is the value of Kaiser-Meyer-Olkin (KMO) statistics which takes a value between 0 and 1. For the application of factor analysis, the

NOTES

NOTES

value of KMO statistics should be greater than 0.5. The KMO statistics compares the magnitude of observed correlation coefficients with the magnitudes of partial correlation coefficients. A small value of KMO shows that correlation between variables cannot be explained by other variables.

Steps in a Factor Analysis Exercise

There are basically two steps that are required in a factor analysis exercise.

1. **Extraction of factors:** The first and the foremost step is to decide on how many factors are to be extracted from the given set of data. This could be accomplished by various methods like the centroid method, the principal component method and the maximum likelihood method. Here, only the principal component method will be discussed very briefly. As we know that factors are linear combinations of the variables which are supposed to be highly correlated, the mathematical form of the same could be written as:

$$F_i = W_{i1}X_1^* + W_{i2}X_2^* + W_{i3}X_3^* + \dots + W_{ik}X_k^*$$

where,

X_i^* = i^{th} standardized variable

F_i = Estimate of i^{th} factor

W_i = Weight or factor score coefficient for i^{th} standardized variable.

k = Number of variables

The principal component methodology involves searching for those values of W_i so that the first factor explains the largest portion of total variance. This is called the first principal factor. This explained variance is then subtracted from the original input matrix so as to yield a residual matrix. A second principal factor is extracted from the residual matrix in a way such that the second factor takes care of most of the residual variance. One point that has to be kept in mind is that the second principal factor has to be statistically independent of the first principal factor. The same principle is then repeated until there is little variance to be explained. Theory may be used to specify how many factors should be extracted or it may be based on the criterion of the Kaiser Guttman method. This method states that the number of factors to be extracted should be equal to the number of factors having an eigenvalue of at least 1. Since each of the variables in the original data set has a variance of 1 (eigenvalue of 1), therefore, if there are 50 variables then the total variation in the data set will be 50.

We know that a factor is a linear combination of the various variables. Now eigenvalue for each of the factor is computed and only those factors that have an eigenvalue at least 1 are accepted as per Kaiser Guttman method. All those factors having eigenvalues less than 1 are rejected. This is because each of the variables has a variance of 1 and, therefore, a linear combination of these variables called factor should not have an eigenvalue less than 1.

Another output of the factor analysis exercise is a *factor score*, which is computed for each of the factors corresponding to each respondent. Most software, including SPSS, provide factor score for each respondent and each factor. As the factor scores are statistically independent, they can be used in regression and discriminant analysis as independent variables. This will be explained briefly in the text later on.

The correlation coefficient of the extracted factor score with a variable is called the *factor loading*. In most computer printouts, a matrix of factor loadings called factor matrix or component matrix is presented. Factor loadings play a very important role in the computations of eigenvalues of each factor and also in computing the communalities of each variable. These concepts would be discussed in depth with the help of a numerical exercise.

2. **Rotation of factors:** The second step in the factor analysis exercise is the rotation of initial factor solutions. This is because the initial factors are very difficult to interpret. Therefore, the initial solution is rotated so as to yield a solution that can be interpreted easily. Most of the computer software would give options for *orthogonal rotation*, *varimax rotation* and *oblique rotation*. Generally, the varimax rotation is used as this results in independent factors. The varimax rotation method maximizes the variance of the loadings within each factor. The variance of the factor is largest when its smallest loading tends towards zero and its largest loading tends towards unity. The basic idea of rotation is to get some factors that have a few variables that correlate high with that factor and some that correlate poorly with that factor. Similarly, there are other factors that correlate high with those variables with which the other factors do not have significant correlation. Therefore, the rotation is carried out in such way so that the factor loadings as in the first step are close to unity or zero. This procedure avoids problems of having factors with all variables having midrange correlations. This is done for a better interpretation of the results and for the ease obtained in naming the factors. Once this is done, a cut-off point on the factor loading is selected. There is no hard and fast rule to decide on the cut-off point. However, generally it is taken to be greater than 0.5. All those variables attached to a factor, once the cut-off point is decided, are used for naming the factors. This is a very subjective procedure and different researchers may name same factors differently. Another point to be noted is that a variable which appears in one factor should not appear in any other factor. This means that a variable should have a high loading only on one factor and a low loading on other factors. If that is not the case, it implies that the question has not been understood properly by the respondent or it may not have been phrased clearly. Another possible cause could be that the respondent may have more than one opinion about a given item (statement).

The total variance explained by all the factors taken together remains the same after rotation. However, the amount of variations for each individual factor

NOTES

may undergo a change. The communalities for each variable under the two procedures remain unchanged.

Cluster

NOTES

Cluster analysis is a grouping technique. The basic assumption underlying the technique is the fact that similarity is based on multiple variables, and the technique attempts to measure the proximity in terms of the study variables. The emerging groups are homogenous in their composition and heterogeneous as compared to the other groups. The grouping can be done for objects, individuals, entities and products. The researcher identifies a set of clustering variables which have been assumed as significant for the purpose of classifying the objects into groups. Thus, it is also referred to as a classification technique, numerical taxonomy and Q analysis. This is basically because the technique is used in various branches of social science, like psychology, sociology, engineering and management. If one were to plot the groups geometrically, a robust cluster analysis is one where individual objects in one cluster are concentrated together and where the individual clusters are far apart from each other. Figure 11.4(a) shows a simple cluster solution of breakfast food based on people who seek nutrition and convenience (ease of preparation).

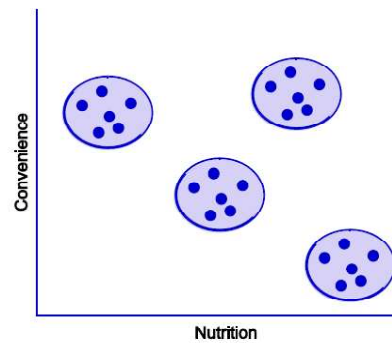


Fig 11.4(a) Ideal Cluster Solution

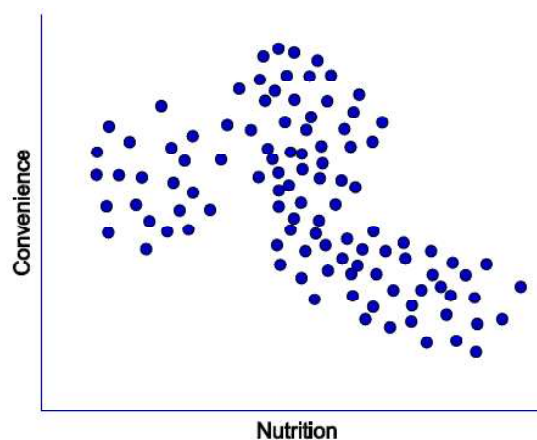


Fig 11.4(b) Actual Cluster Solution

However, the actual situation might be different as the person might be using different criteria for a weekday and for a weekend breakfast. Thus, as the criteria for decision-making become multiple, the grouping does not happen on a simple two-dimensional space but becomes multidimensional [Figure 11.4(b)]. Thus, the researcher is able to group people on these three dimensions and the point regarding the interpretation of benefits sought becomes clear as one understands the multidimensionality of needs. Thus, a bakery/confectionery shop selling sandwiches, patties, bread rolls as well as freshly ground idli batter, using the solution would know: (1) the lucrative segment, (2) the segment which might be motivated to buy if one takes care of their weekday/weekend needs, and (3) A segment which is currently not interested in getting a 'ready-to-eat' breakfast solution and might not look at the bakery as an outlet to visit in the morning. Once the homogenous clusters emerge, the next step is to determine the profile of the group in terms of who they are? What is their gender, age group, family size, etc.? What deals motivate them to buy from a particular store when they are buying eatables in general?

Differentiating Cluster Analysis

In terms of the nature of the technique vis-à-vis the other multivariate techniques, cluster analysis is similar in terms of analyzing the function of multiple independent variables. However, there are essential differences between the other data reduction techniques and cluster analysis.

In factor analysis, the objective was to reduce the original correlated variables to a more manageable number of orthogonal or oblique factors. However, the data reduction was carried out on the columns of the data matrix. On the other hand, in cluster analysis the focus is on the rows, or the individuals or entities and the objective is to group the individuals on the variables.

The other data classification technique is the two group discriminant analysis. Here also, one might wish to group individuals or objects into groups, but the classification or identification of groups is *a priori*. Thus, in the technique one has an established classification rule and the objective of the technique is to validate the information to attest whether the groups obtained by the identified function are correctly classified or not. In cluster analysis, the whole population/sample is undifferentiated and the attempts to assess similarity in response to variables and the grouping happens *post* the clustering.

Usage of Cluster Analysis

Cluster analysis has widespread applicability in all the branches of social sciences and management. In management science, its most valuable contribution is in the area of marketing, especially market segmentation. Some applications of the technique are as follows:

- **Market segmentation:** As we know, Market segmentation is the process of splitting customers/potential customers, within a market into different

NOTES

NOTES

groups/segments, where customers have the same/similar requirement satisfied by a distinct marketing mix (*McDonald and Dunbar*, 1998). This is one area that has seen maximum theorization on the basis of the outputs of the technique. Some examples are **ACORN** (A classification of residential neighbourhood based on 40 variables, e.g., house/car ownership, employment, religion, lifestyle, etc.), **PRIZM** (Potential rating index by zip market. This is based on 39 variables (for example, education, affluence, family life cycle, urbanization, race and ethnicity, mobility, etc.). The solution provides 62 lifestyle categories. The advantage with the technique is that one can look at the combination of variables to predict consumer or potential consumer groups. The best example of clustered solutions are in the area of **benefit segmentation** (*Haley*, 1968). Here, the consumers are divided into groups based on the benefits they seek from the product category. These, then, could be across age groups, gender and other variables. Thus, a marketer could design his product on the basis of this segmentation approach. Yankelovich (1964) segmented consumers in terms of 'what they look for in a watch' and classified people into those who are *price driven*, *durability* and *quality driven*, and those driven by *occasion-bound symbolism*. Sinha (2003) classified food shoppers into *fun* and *work shoppers* based on the benefits they seek from grocery/food purchase. Sondhi and Singhvi (2005) classified grocery shoppers into *transition shoppers*, *traditional shoppers*, *thrifty shoppers* and *indifferent shoppers*.

- **Segmenting industries/sectors:** The researcher could also go about grouping products or sectors (e.g., health or education) into blocks that have some common trait(s). This makes it easier for both the organizations and policy-makers while planning or evaluating the performance of the group.
- **Segmenting markets:** Cities or regions with some common traits like population mix, infrastructure development, climatic or socio-economic conditions could be clustered together. If one city in Kerala and another in Andhra Pradesh are in one cluster, then the organization is able to plan and execute a similar business approach in the two areas.
- **Career planning and training analysis:** In the area of human resources (HR), the technique can be used to group people into clusters on the basis of their educational qualification, experience, aptitude and aspirations. This grouping can assist the HR division to effectively manage training and manpower development for the members of different clusters effectively.
- **Segmenting financial sectors/instruments:** This is an emerging area where different factors like raw material cost, financial allocations, seasonality and other factors are being used to group sectors together to understand the growth and performance of a group of industries. This also assists the policy-makers and the financial analysts in assessing the monetary implications. A number of researchers are making use of clustering principles

to group consumers and their investment behaviour on the basis of the combination of different variables and benefits sought (behavioural finance).

The basic premise of the above technique is, as we said earlier, wherever a researcher wants to manage the data (especially individual or organizational) and he/she perceives that there could be multiple factors involved, cluster analysis is the best classification technique at his/her disposal.

Statistics Associated With Cluster Analysis

Before we review the statistics involved with the technique, it is essential once again to examine the simplicity of the technique. Unlike the other multivariate techniques that we have discussed till now, cluster analysis is the simplest in terms of mathematical derivations. The simplest way to explain the technique is to understand that it simply measures the distance between objects on the basis of multiple variables and looks for similarity as a function of distance, i.e., the shorter the distance between two objects, the more similar they are.

Metric Data Analysis: For obtaining a cluster solution to data that is collected on an interval or ratio scale the statistical assessment of the distance between two objects can be done by calculating the Euclidean distance between them. In case the study has two variables (as stated in the earlier example of nutrition and ease of preparation) then the distance between person A and B can be calculated:

$$d_{A,B} = \sqrt{(X_{B1} - X_{A1})^2 + (X_{B2} - X_{A2})^2}$$

where X_{B1} represents the coordinate of person B on nutrition (interval scale data).

A note of caution here: The Euclidean distance is not 'scale invariant'. It may happen that the relative ordering of the objects in terms of their similarity can be affected by a simple change in the scale by which one or more of the variables are measured. Thus, it is advisable that the data is standardized before being subjected to any analysis. However, it may sometimes happen that standardization can reduce the differences between the groups on the variables that may well be the best discriminators of group differences. Thus, care needs to be taken initially in questionnaire designing to keep the variables measurement scales as roughly of more or less than the same range and avoid standardizing them. Only if the variables are measured on widely different units, standardization is needed to prevent the variables measured in larger units from dominating the cluster solution.

In the example, the two variables were placed on a 10-point scale of importance (with 1 = very important and 10 = very unimportant). The values selected by person A and B were as follows:

Person	Nutrition	Ease of preparation
A	1	2
B	5	2

NOTES

Then the distance between the two is,

$$d_{A,B} = \sqrt{(5-1)^2 + (2-2)^2} = 4.0$$

NOTES

Suppose there was a third person C who had selected

Person	Nutrition	Ease of preparation
C	6	2

Then the distance between A and C would be 5.0 and between B and C would be 1.0.

Thus, B and C are the most similar pair as the inter-person distance is the least and, as stated earlier, the shorter the distance, the greater the similarity.

If, in addition to having nutrition and ease of preparation for breakfast, we also had a variable that measured cost, we would effectively have a 3-dimensional solution. Then the formula would have been:

And generally, for any two objects, i and j:

$$d_{A,B} = \sqrt{(X_{B1} - X_{A1})^2 + (X_{B2} - X_{A2})^2 + (X_{B3} - X_{A3})^2}$$

where,

$$d_{ij} = \sqrt{\sum_{k=1}^n (X_{ik} - X_{jk})^2}$$

d_{ij} = Distance between person i and j

k = Variable (interval/ratio)

i = Object/person

j = Object/person

Also, there are other distance measures available like the city-block or *Manhattan distance* between two objects, which is the sum of the absolute differences in the values for each variable. Another distance measure is the *Chebychev distance* between two objects, which is the maximum absolute difference in values for any variable. However, the most commonly used measure is the squared Euclidean distance. A point to be noted here is that clustering with squared Euclidean distance is faster than the regular Euclidean distance. Thus, for the purpose of clustering, we make use of squared Euclidean distance. The equation for this is the same as the Euclidean distance; only the square root is not calculated.

Then, based on the distance calculated, a distance matrix is created and clusters are created by moving from the most to the least similar pair based on a clustering method.

Multidimensional Scaling (MDS)

The underlying presumptions that one makes while creating an MDS are:

- The individual tries to group objects together.
- The grouped objects are usually evaluated and compared with each other so that they can coexist on a spatial map.
- The basis of evaluation is not unidimensional and the user is at all times (consciously or unconsciously) using an underlying multidimensional space to evaluate the objects.

MDS essentially visually plots the perceptions and preferences of individuals singly and as a group, regarding a group of objects, individuals or both; even when the information about the dimensions or bases of evaluations is minimal.

Thus, the technique uses powerful mathematical tools in order to condense the data by creating visual representations based on the similarities or dissimilarities of data on a spatial map (Schiffman, *et al.* 1981). The map dimensions are hypothesized to be the attributes or features that the person uses to form certain impressions about the object. One of the most widely used mathematical methods to create the maps is based on Kruskal's (1964) stress calculations (to be discussed further in the chapter).

MDS usually involves a comparison of sorts to create a relative position of the considered objects. The comparison could be made on *defined dimensions*, or the apparent basis of comparison. However, more often than not, people make use of their own peculiar and sometimes subjective or *perceived dimensions* to make the comparison. For example, it could be the trust or faith in the service provider in handling the insured person's problems effectively. Thus, two objects or brands with the same defined dimensions might be perceived very differently by the person because:

- The evaluations might not be solely based on defined or observed parameters.
- The subjective and the objective dimensions might be absolutely unrelated.

To simplify the process further, the technique presents the dependent variable (which might be a similarity or dissimilarity between the object or preferences) and then tries to figure out what were the underlying independents or antecedents that led to the obtained map. The advantage of this method is that the researcher's influence where he/she attempts to provide the dimensions of comparison gets minimized. The disadvantage, however, would be to clearly figure out the dimension the respondents might have used for the comparison.

Thus, the researcher needs to be fairly well versed with the probable parameters that a person might use for comparison. These perceived parameters might emerge from a qualitative analysis of the respondents' decision process or through the researcher's review of the secondary literature about the product. The inputs obtained would have to be objectively—without any element of personal

NOTES

NOTES

bias—assessed to comprehend the defined or apparent and the hidden or subjective dimensions being used.

A simple explanation of the concept: To understand the concept of mapping the respondent's choices, let us look at a very simple example of a consumer who buys bread every day for his family breakfast. Now, we ask him which bread he buys. He tells us, 'Harvest Gold, Britannia and Perfect.' Next, we ask him the similarity between two bread brands, say, Harvest Gold and Britannia, on a 7-point scale, where 1 is very similar and 7 is very dissimilar. He says, the similarity is 1. What this means is that:

- If we were to take a mental model of his brain when he said this, the two brands would be very close to each other.
- Suppose we say that the consumer was thinking of price and availability when he was telling us this. Thus, the unconscious evaluation that he did was on the two dimensions of 'price' and 'brand'. So, these two brand are two points close to each other in this two-dimensional map.
- The two manufacturers have to understand that there is no brand loyalty from the customer, as he could very easily buy the competing brand as they are almost identical to each other in his 'mind'.

Now, suppose, we ask him if he has consumed Harvest Gold multi-grain bread, and he says, 'yes'. So we now ask him to tell us the similarity between Harvest Gold regular and Harvest Gold multi-grain bread on the same 7-point scale. His answer is 6. Now, what will happen if we use the same dimensions as in the above case? The brand is the same for both, thus using a two-dimensional map would not be wise as the consumer may be now looking at the health benefit or nutritional content in the breads also as a dimension. Thus this means:

- The bread brands now need a three-dimensional representation to represent their relative positioning in the consumers mind.
- Harvest Gold multi-grain need not worry about competition with the other two as the consumer who buys the multi-grain will not buy them as a substitute as they are very different from the bread they eat regularly.

MDS is only one of the wide array of statistical techniques available for obtaining the object map. The whole range of these methods grouped together is termed as perceptual mapping techniques.

Let us now briefly attempt to understand the underlying algorithms of MDS.

- The inputs obtained by the respondents could be in terms of objects, individuals, brands, corporations or countries.
- The comparison could be in terms of similarities/dissimilarities, e.g. how similar is Delhi to Mumbai on a 7-point scale ranging from the most dissimilar to the most similar; or preferences, e.g. out of the five listed brands, indicate the one you prefer the most to the one that is least preferred.

- As you can observe, the respondent is NOT given any dimension to measure similarity or dissimilarity.
- The preferences could be based on ranked data.
- The respondent might be asked to conduct a paired comparison of the data.

NOTES

Discriminant Analysis

Discriminant analysis is used to predict group membership. This technique is used to classify individuals/objects into one of the alternative groups on the basis of a set of predictor variables. The dependent variable in discriminant analysis is categorical and on a nominal scale, whereas the independent or predictor variables are either interval or ratio scale in nature. When there are two groups (categories) of dependent variable, we have two-group discriminant analysis and when there are more than two groups, it is a case of multiple discriminant analysis. In case of two-group discriminant analysis, there is one discriminant function, whereas in case of multiple discriminant analysis, the number of functions is one less than the number of groups.

Objectives and Uses of Discriminant Analysis

The objectives of discriminant analysis are the following:

- To find a linear combination of variables that discriminate between categories of dependent variable in the best possible manner.
- To find out which independent variables are relatively better in discriminating between groups.
- To determine the statistical significance of the discriminant function and whether any statistical difference exists among groups in terms of predictor variables.
- To develop the procedure for assigning new objects, firms or individuals whose profile but not the group identity are known to one of the two groups.
- To evaluate the accuracy of classification, i.e., the percentage of customers that it is able to classify correctly.

Discriminant analysis can be a very powerful technique of analysis in multiple situations. Some areas in which it is extensively used are as follows:

- **Scale construction:** Discriminant analysis is used to identify the variables/statements that are discriminating and on which people with diverse views will respond differently. For example, in case one wants to assess people who believe that corporate governance is the responsibility of policy-makers against those who think it needs to be self driven or individual centric, one may generate a number of statements and then conduct a pilot study and select only those statements on which the two groups differ significantly.

NOTES

- **Segment discrimination:** Most business managers recognize that the population under consideration can never be totally homogeneous in composition. Therefore, to understand what the key variables are on which two or more groups differ from each other, this technique is extremely useful. Questions to which one may seek answers are as follows:
 - o What are the demographic variables on which potentially successful salesmen and potentially unsuccessful salesmen differ?
 - o What are the variables on which users/non-users of a product can be differentiated?
 - o What are the economic and psychographic variables on which price-sensitive and non-price sensitive customers be differentiated?
 - o What are the variables on which the buyers of local/national brand of a product be differentiated?
- **Perceptual mapping:** The technique is also used extensively to create attribute-based spatial maps of the respondent's mental positioning of brands. The advantage of the technique is that it can present brands or objects and the attributes on the same map. Therefore, the business manager can determine what attribute is the unique selling proposition (USP) of which brand and which are the attributes that are valued by the respondent but there is no brand that currently satisfies that need.

Discriminant Analysis Model

The mathematical form of the discriminant analysis model is:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_K X_K$$

where, Y = Dependent variable

b_s = Coefficients of independent variables

X_s = Predictor or independent variables

It may be kept in mind that the dependent variable Y should be a categorized variable, whereas the independent variables X_s should be continuous. As the dependent variable is a categorized variable, it should be coded as 0, 1 or 1, 2 and 3, similar to the dummy variable coding.

The method of estimating b_s is based on the principle that the ratio of between group sum of squares to within group sum of squares be maximized. This will make the groups differ as much as possible on the values of the discriminant function.

After having estimated the model, the b_s coefficients (also called discriminant coefficient) are used to calculate Y , the discriminant score by substituting the values of X_s in the estimated discriminant model. For any new data point that we want to classify into one of the groups, a decision rule is formulated for this purpose to determine the cut-off score, which is usually the midpoint of the mean discriminant scores of the two groups in case of two-group discriminant analysis, provided the

size of the samples in the two groups are same. The accuracy of classification is determined by using a classification matrix (also called confusion matrix).

The relative importance of the independent variables could be determined from the standardized discriminant function coefficient and the structure matrix. The difference between the standardized and unstandardized discriminant function is that in the un-standardized discriminant function we have a constant term, whereas in the standardized discriminant function, there is no constant term.

NOTES

Check Your Progress

9. What can marketing studies be used for?
10. What is the basic idea of rotation of factors?
11. What is the basic assumption underlying the cluster technique?
12. What is market segmentation?
13. State the advantage and disadvantage of Multidimensional Scaling.
14. List two objectives of discriminant analysis.
15. What is the difference between the standardized and unstandardized discriminant function?

11.9 CONJOINT ANALYSIS

Conjoint analysis uses nominal-scale data. It attempts to identify the most desirable attributes that could be offered in a product or service. An attempt is made to determine the relative importance that consumers attach to the attributes and the utilities that they attach to the levels of attributes. The values assumed by the attributes are called levels. The utilities describe the importance that consumers attach to the levels of each attribute. Here, the respondents are told about the various combinations of the attribute levels and are asked to evaluate the combinations in terms of their desirability. The evaluation can be done either using ordinal or interval-scale data. This will be explained later in the chapter. It may be worth noting that conjoint analysis makes use of subjective evaluation of the combinations presented to the consumer. It makes use of such data to identify the most desirable combinations of the levels of attributes to be included in the new product. In fact, the major business domain where the technique is used is marketing, though it is also applied in the area of HR, Finance and Operations. The various uses of conjoint analysis are to:

- Determine the relative importance of the attributes in the choice process of the consumers
- Determine the market share of brands that differ in attribute levels
- Segment the market based on similarity of preference for attribute levels

NOTES

For conducting the conjoint analysis, the researcher is required to identify the attributes and the levels of the attributes that could be used in constructing the stimuli for presentation to the respondents. The attributes and its various levels could be identified using exploratory research which could be conducted by discussion with management and industry experts; informal interviews with prospective customers, analysis of secondary data and case studies. Once the attributes and its various levels are identified, the respondents are presented with combinations of attributes with levels to show their preference for various combinations. This is illustrated in the following example.

Suppose we ask a set of respondents to express their preference among movies that varied on three attributes, each with two levels as shown below:

- Hero of the movie : Shahrukh Khan or Akshay Kumar
- Type of movie : Action or comedy
- Price of ticket : ₹150 or ₹200

There are in total $2 \times 2 \times 2 = 8$ combinations of these features. Each of these features is presented to, say, respondent number 1. The various features would look like:

Feature 1 – Shahrukh Khan, Action, ₹150

Feature 2 – Shahrukh Khan, Action, ₹200

Feature 3 – Akshay Kumar, Action, ₹150

Feature 4 – Akshay Kumar, Action, ₹200

Feature 5 – Shahrukh Khan, Comedy, ₹150

Feature 6 – Shahrukh Khan, Comedy, ₹200

Feature 7 – Akshay Kumar, Comedy, ₹150

Feature 8 – Akshay Kumar, Comedy, ₹200

The respondent could be presented with the above eight combinations and asked to give their preferences in terms of desirability of the feature, either on an interval scale or ordinal scale.

Steps in Conjoint Analysis

The following steps are involved in carrying out a conjoint analysis exercise.

1. Identification of Attributes

As a first step, the researcher needs to identify the various attributes that may be used in constructing stimuli. It is important from the point of view of both the consumer and the company. From the consumer point of view, only those attributes that influence the consumers' choice will be selected. This is determined through exploratory research, for example, through managerial judgments. From the point of view of the company, it gains importance because the company has to see

whether it has the technological or other resources which could be used to incorporate consumer preferences. One has to be careful in selecting the attributes since only a limited number could be used in a conjoint study.

2. Determination of Attribute Levels

This involves specifying the actual levels for each attribute. In a study on the sale of juices, if the attribute of interest is flavour, then the possible corresponding levels could be mango, orange, and pineapple. The number of levels of each attribute has a direct bearing on the number of stimuli which the respondents could be assumed to evaluate. To ensure the quality of data, it is advisable to have as small number of levels for each attribute as possible. The goal is to end up with good estimates of the utility of each attribute level.

While creating stimuli for conjoint judgment task, the researcher has to keep in mind that there is a relationship between the number of levels used to measure an attribute and the inferred importance of the attribute. Empirical evidence suggests that the more the levels used for an attribute, the more important that attribute is going to be inferred in the analysis.

The choice of attribute levels depends upon their effect on consumer choice. If the chosen attribute levels lie outside the range that is usually encountered, it decreases the believability by respondents but could increase the accuracy by which parameters can be estimated. A combination such as very low price and high quality is unbelievable, but it could result in increase in accuracy by which parameters can be estimated.

3. Determination of Attribute Combinations

In the third stage, the specific combination of various attributes is decided upon. If there are three attributes each with three levels, the total number of stimuli to be presented to the respondents would be $3 \times 3 \times 3 = 27$. Suppose each of the attributes has 5 levels—this would result in $5 \times 5 \times 5 = 125$ stimuli. It will certainly be very difficult for respondents to provide meaningful judgments on their preference of such a large number of stimuli. The analyst, instead of using the full factorial design where the respondent is to indicate preference for all stimuli, uses a fractional factorial design where only select stimuli are selected.

4. Nature of Judgment on Stimuli

The next step is to ascertain the nature of judgment on stimuli. This can be done in two ways. One is to ask the respondents to rank order the various stimuli according to their preference or intention to buy. The advantages of ranking are their ease of use by consumers, ease of administration, etc. The second way is to use a rating scale. The respondent may be asked to rate their preference on a 5-point or 7-point rating scale. The advantages of using rating scale are that it is less time consuming, more convenient to use and easier to analyse. In the rank-order

NOTES

approach, relative judgments are sought, whereas in rating scale, respondents are asked to indicate their degree of preference for each stimuli.

5. Aggregation of Judgments

NOTES

The fifth step is to decide how the responses from various individual consumers are aggregated. One option is to estimate the utility function for each individual. The problem with such an analysis is that individual-level functions cannot be used for formulating marketing strategies. On the other extreme, one could pool the results across all respondents and estimate one overall utility function. This approach ignores the heterogeneity that may exist among respondents. The best option would be to group respondents in the form of segments. This will have clear marketing strategy implications for managers. The main question, however, is how to form segments. The segments formed are homogenous with respect to the benefits that respondents want from product or service.

6. Choice of Technique of Analysis

The last step would be to choose the technique so that the input data can be analysed. As our purpose is to estimate the utilities for each level of each attribute, a dummy variable regression would be quite appropriate.

Illustration of Conjoint Analysis with an Example

As another example, consider the case of a soft drink bottling company that wants to enter the fruit juice market, as the demand for aerated drinks is showing a stagnant trend because of increasing health consciousness among the consumers. Before the launch of fruit juices, the company wants to undertake a study to determine what combination of attributes with various levels would be the most desired one. The three attributes considered in the study are flavour, packaging and price each of them having three levels (Table 11.18).

Table 11.18 Fruit juice attributes and levels

Attribute	Number	Level Description
Flavour	3	Mixed Fruit
	2	Orange
	1	Mango
Packaging	3	Tetra pack
	2	Plastic Bottle
	1	Glass Bottle
Price	3	₹90/-
	2	₹75/-
	1	₹65/-

Two approaches are used for constructing conjoint analysis stimuli: the pair-wise approach and full profile approach. The full profile approach lists all the stimuli in terms of all attributes by using the attribute levels specified by the design. This chapter does not make use of the pair-wise approach. The full profile approach is a multiple factor evaluation and is being used in the present example. In this approach, complete profiles are considered for all the attributes. Each profile is described on a card and respondent is asked to evaluate the same in terms of its preference on a 9 – point interval scale where 1 = least preferred to 9 = most preferred. Given three attributes, defined at three levels each, a total of $3 \times 3 \times 3 = 27$ profiles can be constructed.

In order to reduce the task of respondent evaluation, a fractional factorial design is employed and the set of nine profiles is constructed. The purpose of fractional factorial design is to reduce the number of stimuli profile to be evaluated out of the full profile. In the present example, the set of nine profiles is constructed, which constitutes the estimation stimuli (Table 11.19).

Table 11.19 Fruit juice profiles and their ratings

Profile No.	Flavour	Packaging	Price	Preference Rating
1	1	1	1	5
2	1	2	2	6
3	1	3	3	5
4	2	1	2	7
5	2	2	3	7
6	2	3	1	9
7	3	1	3	4
8	3	2	1	8
9	3	3	2	7

Conjoint analysis makes use of dummy variable regression, where dependent variable preference is treated as utility or the Part-worth of the level of attributes. The utility model is written as:

$$U = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + U$$

where, X_1, X_2 = dummy variables representing flavour
 X_3, X_4 = dummy variables representing packaging
 X_5, X_6 = dummy variables representing price

where, X_1 = 1, if choice is for mango
 = 0, Otherwise
 X_2 = 1, if choice is for orange
 = 0, Otherwise

NOTES

NOTES

$X_3 = 1$, if choice is for glass bottle
 $= 0$, Otherwise

$X_4 = 1$, if choice is for plastic bottle
 $= 0$, Otherwise

$X_5 = 1$, if choice is for ₹65/-
 $= 0$, Otherwise

$X_6 = 1$, if choice is for ₹75/-
 $= 0$, Otherwise

U = Preference rating or utility

We have used level 3 as the base level. The data for respondent number 1 is presented in Table 11.20.

Table 11.20 Fruit juice data for dummy variable regression

S. No.	X_1	X_2	X_3	X_4	X_5	X_6	Y
1	1	0	1	0	1	0	5
2	1	0	0	1	0	1	6
3	1	0	0	0	0	0	5
4	0	1	1	0	0	1	7
5	0	1	0	1	0	0	7
6	0	1	0	0	1	0	9
7	0	0	1	0	0	0	4
8	0	0	0	1	1	0	8
9	0	0	0	0	0	1	7

The estimated regression equation is given below:

$$U = 5.778 - 1.00X_1 + 1.333X_2 - 1.667X_3 + 0.000X_4 + 2.000X_5 + 1.333X_6$$

Here,

$$b_0 = 5.778$$

$$b_1 = -1.00$$

$$b_2 = 1.333$$

$$b_3 = -1.667$$

$$b_4 = 0.000$$

$$b_5 = 2.000$$

$$b_6 = 1.333$$

U stands for utility.

Each dummy variable coefficient represents the difference in the part-worth for that level minus the part-worth for the base level. For flavour, we have the following:

$$\alpha_{11} - \alpha_{13} = b_1$$

$$\alpha_{12} - \alpha_{13} = b_2$$

An additional constraint is required since the part-worths are estimated on an interval scale, which has an arbitrary origin. The additional constraint looks like:

$$\alpha_{11} + \alpha_{12} + \alpha_{13} = 0$$

The equations for fruit juice are:

$$\alpha_{11} - \alpha_{13} = -1.00$$

$$\alpha_{12} - \alpha_{13} = 1.333$$

$$\alpha_{11} + \alpha_{12} + \alpha_{13} = 0$$

Solving these equations, we get

$$\alpha_{13} = -0.111$$

$$\alpha_{12} = 1.333 - 0.111$$

$$= 1.222$$

$$\alpha_{11} = -1.111$$

The equations for second attribute (packaging) are:

$$\alpha_{21} - \alpha_{23} = b_3$$

$$\alpha_{22} - \alpha_{23} = b_4$$

$$\alpha_{21} + \alpha_{22} + \alpha_{23} = 0$$

$$\alpha_{21} - \alpha_{23} = -1.667$$

$$\alpha_{22} - \alpha_{23} = 0$$

$$\alpha_{21} + \alpha_{22} + \alpha_{23} = 0$$

$$\therefore \alpha_{23} = 0.556$$

$$\alpha_{21} = -1.111$$

$$\alpha_{22} = 0.556$$

Similarly, for the third attribute (price), we have

$$\alpha_{31} - \alpha_{33} = b_5$$

$$\alpha_{32} - \alpha_{33} = b_6$$

$$\alpha_{31} + \alpha_{32} + \alpha_{33} = 0$$

$$\alpha_{31} - \alpha_{33} = 2.000$$

$$\alpha_{32} - \alpha_{33} = 1.333$$

NOTES

$$\alpha_{31} + \alpha_{32} + \alpha_{33} = 0$$

$$\alpha_{33} = -1.111$$

$$\alpha_{31} = 0.889$$

$$\alpha_{32} = 0.222$$

NOTES

The relative importance of attributes indicates which attributes are important in influencing the choice of the consumers. The relative importance weights are calculated based on ranges of part-worths as follows:

$$\begin{aligned} \text{Sum of ranges of part-worths} &= [1.222 - (-1.111)] + [0.556 - (-1.111)] + (0.889 - (-1.111)) \\ &= 2.333 + 1.667 + 2.0 \\ &= 6.000 \end{aligned}$$

$$\text{Relative importance of flavour} = \frac{[1.222 - (-1.111)]}{6.0} = \frac{2.333}{6.0} = 0.39$$

$$\text{Relative importance of packaging} = \frac{[0.556 - (-1.111)]}{6.0} = \frac{1.667}{6.0} = 0.28$$

$$\text{Relative importance of price} = \frac{[0.889 - (-1.111)]}{6.0} = \frac{2.0}{6.0} = 0.33$$

The results for part-worths and relative contribution of attributes are given in Table 4.47:

The estimation of part-worths and the relative importance of weights provide the basis for interpreting the results. In the case of our respondent, the weight assigned by him to flavour, price and packaging are 39, 33 and 28 per cent respectively. It is seen that the respondent prefers orange flavour, followed by mixed fruit and mango. The respondent is indifferent between plastic and tetra packaging. As expected, the price of ₹65 has the highest utility and the price of ₹90 has the lowest utility.

Table 11.21 Results of conjoint analysis

Attribute	Number	Description	Utility	Importance
Flavour	3	Mixed fruit	-0.111	0.39
	2	Orange	1.222	
	1	Mango	-1.111	
Packaging	3	Tetra pack	0.556	0.28
	2	Plastic bottle	0.556	
	1	Glass bottle	-1.111	

The results can be interpreted better by plotting the part-worth function in Figure 11.5, which is self-explanatory.

Uses of Conjoint Analysis

The main uses of Conjoint Analysis are as follows:

- **Segmentation:** Cluster analysis is a technique that could be used to segment the respondents of conjoint analysis exercise. The segmentation exercise could be based on the similarities/dissimilarities in the utilities that are attached to the levels of various attributes. The analysis could group customers of two-wheelers that care about different attributes—there could be respondents to whom product features are important, some for whom economic considerations are important, while others may attach importance to pride of ownership.

NOTES

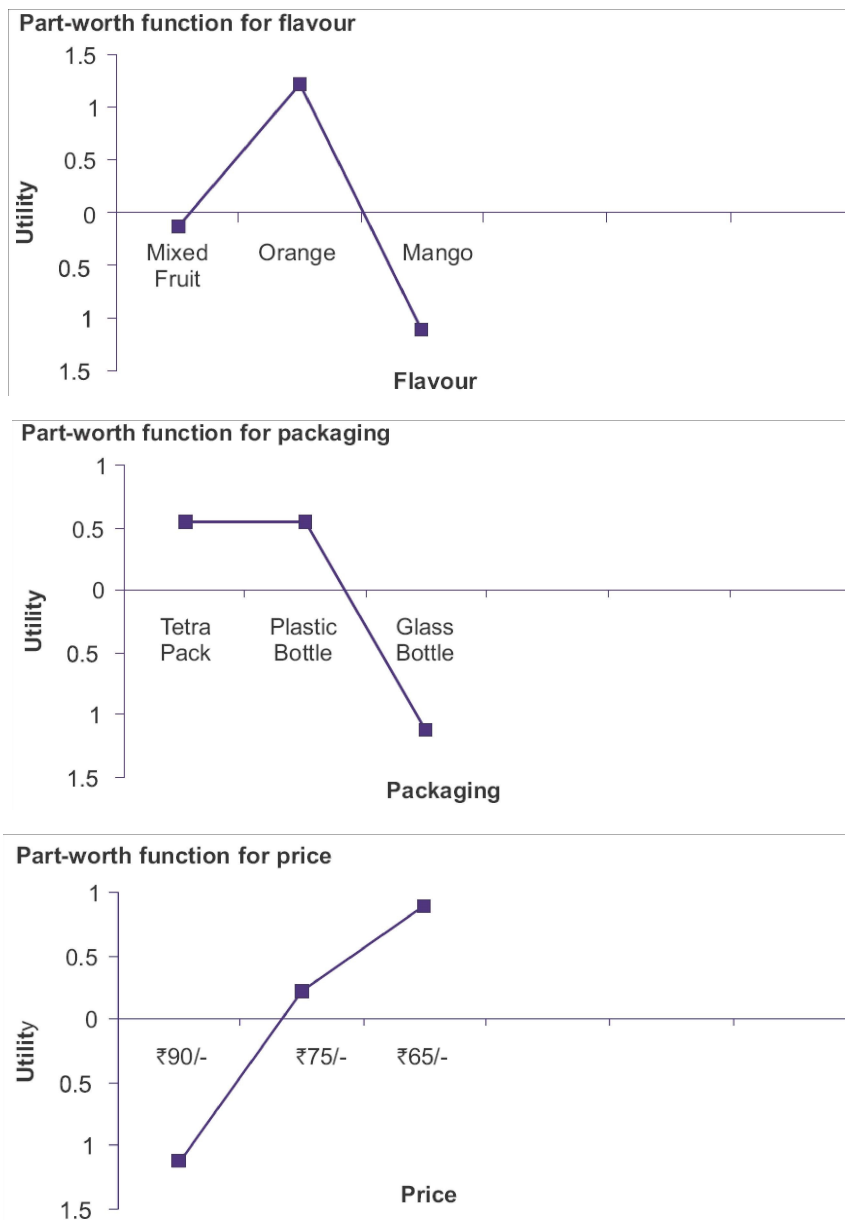


Fig. 11.5 Part-worth functions

NOTES

- **Computation of price elasticity:** In case price is included in the conjoint analysis, one of the outputs obtained could be a utility function for price, which could be used for computation of elasticity. Elasticity information can be obtained at a lower cost from such data rather than from conjoint analysis. To get elasticity information for a mixer grinder, describe the product to the respondents and ask whether they would be willing to pay 'X' for it. Divide the respondents into several groups, say, five to six, and give a different value of X. Develop a rough demand curve by plotting the percentage of respondents who would buy at each price. This would provide information about price elasticity at lesser cost than a conjoint procedure. Therefore, the conjoint analysis should not be used for the sole purpose of computing price elasticity.
- **Estimating sales for new or improved products:** Conjoint analysis could be used for estimating sales for new or improved products. The effect of changes in any attribute like price, product features, warranty terms in the conjoint procedure can help in answering “what if” this occurred.

Check Your Progress

16. What does conjoint analysis attempt to do?
17. What happens if the chosen attribute levels lie outside the range?

11.10 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The common ways of summarizing data are by calculating average, range, standard deviation, frequency and percentage distribution.
2. The type of descriptive analysis to be carried out depends on the measurement of variables into four forms—nominal, ordinal, interval and ratio.
3. Frequency distribution is the counting of responses or observations for each of the categories or codes assigned to a variable.
4. Regression is used to explain the variations in one variable—usually called the dependent variable—by a set of independent variables. It identifies the nature of the relationship.
5. The basic principle underlying the technique is that the total variation in the dependent variable is broken into two parts—one which can be attributed to some specific causes and the other that may be attributed to chance.

6. The main advantage of factorial design over randomized block design is that it is possible to measure the main effects as well as the interaction effects of two or more independent variables at various levels.
7. Analysis of co-variance represents an extension of analysis of variance that tests the significance of difference between means of final experimental data by taking into account the correlation between the dependent variable and one or more co-variates or pertinent control variables, and by adjusting initial mean differences in the groups.
8. The assumption of homogeneity of sample variances is tested by applying Bartlett's test of homogeneity of variance.
9. Marketing studies can be successfully used for new product development; product acceptance research, developing of advertising copy, pricing studies and for branding studies.
10. The basic idea of rotation is to get some factors that have a few variables that correlate high with that factor and some that correlate poorly with that factor.
11. The basic assumption underlying the technique is the fact that similarity is based on multiple variables, and the technique attempts to measure the proximity in terms of the study variables.
12. Market segmentation is the process of splitting customers/potential customers, within a market into different groups/segments, where customers have the same/similar requirement satisfied by a distinct marketing mix.
13. The advantage of this method is that the researcher's influence where he/she attempts to provide the dimensions of comparison gets minimized. The disadvantage, however, would be to clearly figure out the dimension the respondents might have used for the comparison.
14. The objectives of discriminant analysis are the following:
 - To find a linear combination of variables that discriminate between categories of dependent variable in the best possible manner.
 - To find out which independent variables are relatively better in discriminating between groups.
15. The difference between the standardized and unstandardized discriminant function is that in the un-standardized discriminant function we have a constant term, whereas in the standardized discriminant function, there is no constant term.
16. Conjoint analysis attempts to identify the most desirable attributes that could be offered in a product or service. An attempt is made to determine the

NOTES

NOTES

relative importance that consumers attach to the attributes and the utilities that they attach to the levels of attributes.

17. If the chosen attribute levels lie outside the range that is usually encountered, it decreases the believability by respondents but could increase the accuracy by which parameters can be estimated.

11.11 SUMMARY

- Descriptive analysis refers to transformation of raw data into a form that will facilitate easy understanding and interpretation. Descriptive analysis deals with summary measures relating to the sample data.
- Under inferential statistics, inferences are drawn on population parameters based on sample results. The researcher tries to generalize the results to the population based on sample results.
- The first step under univariate analysis is the preparation of frequency distributions of each variable.
- In cross-tabulation, responses to two questions are combined and data is tabulated together. A cross-tabulation counts the number of observations in each cross-category of two variables.
- Simple correlation measures the degree of association between two variables. The correlation could be positive, negative or zero.
- Analysis of Variance (ANOVA) helps in performing this test in one go and, therefore, is considered to be important technique of analysis for the researcher.
- In ANOVA, the dependent variable in question is metric (interval or ratio scale), whereas the independent variables are categorical (nominal scale).
- Completely randomized design involves the testing of the equality of means of two or more groups. In this design, there is one dependent variable and one independent variable.
- The variation within the sample, which is attributed to chance, is referred to as the error sum of squares (SSE). This could be computed by subtracting the treatment sum of squares from the total sum of squares.
- In factorial design, the dependent variable is the interval or the ratio scale and there are two or more independent variables which are nominal scale.
- Analysis of covariance is useful to experimental psychologist, especially in the field of memory and learning, when for various reasons it is not possible or quite difficult to equate experimental groups on some pertinent variable at the start of the experiment.

- The assumption that there should be a real correlation between the dependent variable and the pertinent control variable is tested by using the *Test of Significance of Regression*.
- Factor analysis is a multivariate statistical technique in which there is no distinction between dependent and independent variables. In factor analysis, all variables under investigation are analysed together to extract the underlined factors. Factor analysis is a data reduction method.
- Factor analysis could be used to develop concise multiple item scales for measuring various constructs.
- Establishing antecedents reduces multiple input variables into grouped factors. Thus, the independent variables can be grouped into broad factors.
- Factor analysis exercise requires metric data. This means the data should be either interval or ratio scale in nature.
- The first and the foremost step in factor analysis is to decide on how many factors are to be extracted from the given set of data. This could be accomplished by various methods like the centroid method, the principal component method and the maximum likelihood method.
- The second step in the factor analysis exercise is the rotation of initial factor solutions. This is because the initial factors are very difficult to interpret.
- Cluster analysis is a grouping technique. The basic assumption underlying the technique is the fact that similarity is based on multiple variables, and the technique attempts to measure the proximity in terms of the study variables.
- In factor analysis, the objective was to reduce the original correlated variables to a more manageable number of orthogonal or oblique factors. However, the data reduction was carried out on the columns of the data matrix.
- Cluster analysis has widespread applicability in all the branches of social sciences and management. In management science, its most valuable contribution is in the area of marketing, especially market segmentation.
- In the area of human resources (HR) the technique can be used to group people into clusters on the basis of their educational qualification, experience, aptitude and aspirations.
- MDS essentially visually plots the perceptions and preferences of individuals singly and as a group, regarding a group of objects, individuals or both;

NOTES

NOTES

even when the information about the dimensions or bases of evaluations is minimal.

- Discriminant analysis is used to predict group membership. This technique is used to classify individuals/objects into one of the alternative groups on the basis of a set of predictor variables.
- The values assumed by the attributes in a conjoint analysis are called levels. The utilities describe the importance that consumer attach to the levels of each attribute.
- For conducting the conjoint analysis, the researcher is required to identify the attributes and the levels of the attributes that could be used in constructing the stimuli for presentation to the respondents.
- While creating stimuli for conjoint judgment task, the researcher has to keep in mind that there is a relationship between the number of levels used to measure an attribute and the inferred importance of the attribute.
- In order to reduce the task of respondent evaluation, a fractional factorial design is employed and the set of nine profiles is constructed. The purpose of fractional factorial design is to reduce the number of stimuli profile to be evaluated out of the full profile.
- Cluster analysis is a technique that could be used to segment the respondents of conjoint analysis exercise. The segmentation exercise could be based on the similarities/dissimilarities in the utilities that are attached to the levels of various attributes.

11.12 KEY WORDS

- **Descriptive Analysis:** It refers to transformation of raw data into a form that will facilitate easy understanding and interpretation. It deals with summary measures relating to the sample data.
- **Analysis of Variance (ANOVA):** It is a collection of statistical models and their associated estimation procedures used to analyze the differences among group means in a sample. ANOVA was developed by the statistician Ronald Fisher.
- **Factor Analysis:** It is a multivariate statistical technique in which there is no distinction between dependent and independent variables.
- **Factor Score:** It is a numerical value that indicates a person's relative spacing or standing on a latent factor.

- **Multidimensional Scaling (MDS):** It is a means of visualizing the level of similarity of individual cases of a dataset. MDS is used to translate “information about the pairwise ‘distances’ among a set of n objects or individuals” into a configuration of n points mapped into an abstract Cartesian space.
- **Cluster Analysis:** It is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.

NOTES

11.13 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write a short note on cross-tabulation.
2. List the two assumptions made in the ANOVA technique.
3. What is error sum of squares (SSE)?
4. What is a factor?
5. Write a short note on Kaiser Guttman method.
6. How do factor analysis and cluster analysis differ?
7. What are the underlying presumptions that one makes while creating an MDS?
8. List the underlying assumptions of MDS.
9. What is discriminant analysis used for?
10. Write a short note on perpetual mapping?

Long-Answer Questions

1. Elaborate upon the three types of correlation.
2. Discuss the randomized block design in two-way ANOVA.
3. Explain two ways of testing basic assumptions underlying the technique of analysis of co-variance.
4. Analyze the uses of factor analysis.
5. Describe the conditions for a factor analysis exercise.
6. Discuss the areas in which discriminant analysis is used.
7. Explain the steps involved in conjoint analysis.

NOTES

11.14 FURTHER READINGS

- Saravanavel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.
- Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.
- Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.
- Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.
- Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

BLOCK - IV
REPORT WRITING AND RECENT TREND
IN MARKETING RESEARCH

*Data Interpretation and
Report Writing*

NOTES

**UNIT 12 DATA INTERPRETATION
AND REPORT WRITING**

Structure

- 12.0 Introduction
- 12.1 Objectives
- 12.2 Data Interpretation
- 12.3 Research Report
 - 12.3.1 Modus Operandi of Writing a Market Research Report
 - 12.3.2 Structure and Components of the Report
 - 12.3.3 Style, Layout of a Report and Finalizing the Report
 - 12.3.4 Quality Research Report
- 12.4 Responsibilities of a Market Research Report Writer
- 12.5 Presenting the Report
- 12.6 Answers to Check Your Progress Questions
- 12.7 Summary
- 12.8 Key Words
- 12.9 Self Assessment Questions and Exercises
- 12.10 Further Readings

12.0 INTRODUCTION

A research report is a recorded data or document that outlines the processes and results of an investigation or survey. The main functions of a research report are to communicate the results of the marketing research, to help executives in the selection of right action, to verify the conclusions of other reports and so on. For a report to be comprehensible and fruitful, it is essential that it is presented in a systematic and lucid manner. A report should be clear, concise, easy to understand, objective and accurate so that the objective of its formulation is met. This unit provides an overview of data interpretation and the dimensions of research report writing.

12.1 OBJECTIVES

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

- Describe the Modus Operandi of writing a Research Report
- Analyse the structure, components and quality of a Research Report

- Discuss the style, layout of a report and finalizing the report
- Explain the responsibilities of a Market Research Report writer and the ways of presenting the report

NOTES

12.2 DATA INTERPRETATION

The researcher, once the statistical analysis is over, proceeds the next and important measure termed *interpretation*. Interpretation of research results occurs at the beginning, the researcher may summarise and bring together the various numerical values to the forefront. This will enable easy comparison and comprehension. Generally, the readers do not like to probe into the details of calculations of statistical value, but likely to see the final findings. This means summarising the calculated values properly in a feasible manner is very essential. Accepting or rejecting the null hypothesis is to end with the statistical analysis and interpretation.

The researcher has to draw the real cause and reasons thereof for accepting or rejecting the hypothesis based on the methodology approach, conditions of study and so on. This interpretative writing is key step. The beginner of research needs guidance. He should approach with an open mind and keep off ideological and personal biases from polluting the research. Being continually conscious of these shortcomings is the way of improving the quality of interpretative and innovative writing the research report.

Meaning of Interpretation

Interpretation of research (statistical) results means explaining the figures and facts in the context of theory on which the investigation is done. Thus, interpretation serves a dual purpose. First, it gives an understanding of the general factors that seem to explain what has been observed and second, it provides theoretical conception which can serve as a guide for further research.

The task of researcher is incomplete if he fails by presenting his findings of investigation that arrived at through the analysis of data. For example, the researcher winds up his investigation just by saying that children have a greater propensity to have more clothes compared with others is hardly meeting his obligation to research. He, in the large interest of research, must also show his observations to the underlying relations and processes which are initially hidden to eye. To refer to our example, the researcher should be able to show his observation that the children have a greater propensity to have clothes on account of deeper relationship between parents love and children joy, which would structure good relations between them.

In essence, the researcher's work goes well beyond the collection and analysis of data. His task extends to concluding the findings of his investigation. It is through interpretation that the researcher can understand the real significance and know the real purpose of his findings. He appreciates why the findings are? What they are? Interpretation is intertwined with analysis. But, it is better to conceive as

special aspect of analysis rather than a separate or distinct operation. In brief, interpretation is search for the meaning of research findings. It serves to open new and more widely avenues of intellectual exercise and promotes the interest for new knowledge and for greater wisdom.

Precautions in Interpretations

It is important to recognise the errors crept in interpretation. To guard this, one is to be careful and critical of own thinking and thereon making satisfactory interpretations. While interpreting the statistical results, most possible committed errors are :

1. The researcher, sometimes, lacks of adequate grasp and insight of the problem in broad sense keeping the future.
2. The relevance of various elements should be interpreted. If fails, it may cause of overlook the operations of significant factors.
3. Non-responsiveness in sampling, bias in data, inadequacy in design, defective data collection tools, inaccurate statistical analysis, etc., may be well recognised.
4. The unstudied factors are taken into account, it may lead in wrong interpretation. The research result is composed of many factors but does not produce by a single factor. Average of those factors which have not studied if taken the same for each factor under an assumption, would provide faulty findings. This assumption, no doubt, is convenient but not safe as well as sound.
5. Proper interpretation of data lies on proper evaluation of facts. For example, if opinion relating to ONIDA is studied and 50 percent of respondents express the BPL is satisfactory while 50 per cent not. What the researcher understands by this data? He is not in a position to judge which of the two groups is right unless he studied and evaluated the ONIDA performance thoroughly by himself.

Fallacies

In interpreting the investigator has to try to avoid the common fallacies (mean really illogical) but apparently genuine for arriving at the right conclusions. The common fallacies are:

- Argument from a single or very limited number of facts.
- Argument from positive instances neglecting negative points.
- Argument from analogy.
- Omission of factor evidence to one's opinion.
- Failure to note important facts attending different phenomena.
- Attributing to a single variable where two or more variables results require.

NOTES

NOTES

- Wrong conclusions due to preconceived ideas and prejudices.
- Inadequate tools of measurement.
- Depending upon subjective judgement and guesses.
- Errors incidental to classification, nomenclature and terminology.
- Faulty or mal-observation.
- Generalisations from insufficient data.

Measures to Be Followed

While interpreting the statistical results and getting generalisations, the following measures should be kept in mind. They are :

1. The conclusions should be on the evidence of sound and adequate data.
2. The conclusions should be drawn through the use of right technique in the right context.
3. The conclusions and generalisations must meet the questions asked in the statement of problem.
4. The conclusions must prove or disprove the hypothesis formulated for the purpose of study at the beginning.
5. The conclusions and generalisations must be accompanied by recommendations for application and implementation.

12.3 RESEARCH REPORT

The effectiveness of the research depends upon the methods of communication and presentation of the research report. A very useful research, if not presented in a suitable manner for the users of the research findings, may not serve its purpose. The users are not greatly interested in the research methodology, they will use only those findings which greatly interested them. They use only those findings that are easy to understand and can be used to solve their problems. Therefore, it is very essential to present the research findings in a very effective manner to meet the requirements of the marketing managers. It is accepted that properly conducted research may be presented effectively. The researcher may put the objectives of the research first, followed by suggestions and how to solve the problems or methods to attain the objectives. Properly conducted research can be presented effectively, but a haphazard manner of conducting a research may result in a failure. An oral research report may not be properly understood. Therefore, it is essential to present the report in black and white. The presentation of research findings can be studied under the objective of the report, types of report, content of the report, principles of report presentation, visual devices in report preparation, presentation and communication methods and proper follow-up.

Report writing is the final phase of a research investigation. After the collected data has been analyzed, various conclusions and generalizations have been drawn

through interpretation, after which the report has to be prepared. The report of research, thus, is a statement of the contents in brief, the procedure adopted and findings arrived at by the researcher. It is not a complete description of the activities carried out during the period of investigation.

Reporting research results may appear simple, but it requires considerable thought, artistic skill, effort, patience and penetration, covering all approach to the problem and data analysis, control over language, and greater philosophy.

12.3.1 Modus Operandi of Writing a Market Research Report

The purpose of a report is to reveal to the interested readers the problem investigated, the methods used to solve the problem, the results of research and the conclusions inferred. A research would remain incomplete if it is not given the shape of a report and presented to the audience. The American Marketing Society has defined the research report as;

‘To convey to the interested persons the whole results of the study in sufficient detail, so arranged as to enable each reader to comprehend the data and to determine for himself the validity of conclusions.’

It means the purpose of report is to invite others to study the generalizations and offer their own comments regarding them.

Research is primarily a cooperative venture. It is essential that every researcher should know what others have found in their studies. Thus, it involves dissipation of knowledge and also creates ground for hypothesis and leads to further research on the same or an allied problem. By means of a report, various small segments of research can be coordinated and consolidated into a single theory. Coordination is not possible unless a complete report is available concerning the procedure adopted and the findings identified in different researches.

Classification of Reports

According to the purpose of the report, it can be classified as a basic report, a report for publication, a technical report and a report intended for executives.

- 1. Basic report:** The basic report is the first report prepared for the research project. It is prepared by the researcher for his own use as a working paper for further study and findings. Only experienced researchers are accustomed to prepare such a report as a qualitative final report. Many of the researchers do not bother to prepare such reports. They prepare the final report, which may be full of preliminary shortcomings. Unless basic reports are prepared and complete basic records are maintained, final findings may not incorporate the necessary data and information. The final report may not be very useful. Therefore, it is essential to prepare the basic report from time to time to record all available conclusions and findings.

NOTES

NOTES

2. Reports for publication: During the course of research, the researchers may prepare research papers suitable for publication in magazines, journals and newsletters. They may provide an insight into a problem. They get a feedback from the public and team experts. The research paper does not form part of the final report as it is generally written, keeping in view the interests of the public and the outlook of journals. Research papers are published only in technical and concerned periodicals. The papers written for readers of general interest will have to be moulded in the language of the common people. The research papers have proved to be very useful for the public, for new researchers and management. In India, many important journals in the areas of marketing, finance, industrial relations and management have been publishing research papers for the benefit of business, industry and professionals.

3. Technical report: Technical reports are prepared by specialized people, such as technically trained personnel and academicians. Such people prepare the reports in a logical sequence, using relevant statistics and comprehensive data as appendices. The problems and hypotheses are properly researched and their findings are given in technical terms. The sources of the data are properly acknowledged. Appendices and diagrams are given to substantiate the findings and conclusions.

4. Reports for executives: These reports are the actual and final reports for the decision makers in the areas of marketing management. The researchers prepare the report keeping in view the objectives of the management. The reports for executives suggest the appropriate measures to be adopted by the management to solve problems or arrive at concrete decisions in management.

Functions of Research Report

A market research report performs the following functions:

- To communicate the marketing research result
- To help the executives in selection of right action
- To verify the conclusions of other reports
- To provide an understanding of data and conclusions thereof
- To aid the executive in judgement of quality of research work
- To serve as an essential reference for future research

Types of Reports

While drafting the research report, one should keep in mind the level of the knowledge of the readers. A report written for experts or scientists of marketing institutions must be more vigorous, more detailed than what is written for the general public. Before the preparation of the report, the researcher has to consider the

background of potential audience, their situation, their technical profile and their knowledge. Reports are of three types. They are as follows:

- (i) Report to the layman
- (ii) Report to the administrator
- (iii) Report to the technical expert

The report to the layman is also termed as the popular report. It is meant for the layman with a view to disseminate the broad facts, findings and recommendations of the investigation. Writing for the layman is more time consuming than writing for a fellow professional. The reason being that the use of a simple, yet dignified terminology, without the use of technical jargon is quite tricky. The researcher has to take special efforts to put his scientific knowledge before the layman in a form that could be understood by him.

The report for the administrator should neither be too general nor too technical. However, it should be of medium size with little technical terms and jargons, supporting data followed by a summary and the major recommendations. At times, this helps the management in its policy decision-making.

The technical report is meant for the fellow professional / researcher, but not for the layman. This type of report should naturally contain all the details of the research study. Hence, it will be more technical. Research reports may be in form of a thesis, dissertation, monograph, journal-article, sponsored research project and so on. Due to space constraints and selectivity, a journal-article cannot discuss every aspect of research in detail as can be done in a thesis. Before preparing a paper, the researcher should refer to the style sheet of the concerned journal.

The thesis writing is to be dealt in detail as well as in depth on specific aspects undertaken for investigation. Concrete conclusions and thereupon, recommendations should be made for the well-being of the society at large. However, it may not be the same style everywhere, but, by and large, a pattern emerges. Science of the drafting style of thesis involves technical phrases and is specifically meant for its appraisal, it is also termed as a *technical report*.

In the research report, there should be a brief statement concerning the authorization of the study, the problem to be examined and its breadth and depth, following this, a summary along with conclusions and recommendations are included. Short reports are appropriate for studies in which the problem is well-defined within the scope. They need only limited time, personnel and money for which methodological issues are simple and straight-forward.

Before the presentation of the market research data, every researcher should pay heed to the following aspects:

- *Produce honest and sober interpretations.* A market research report need not be a bombshell. If the findings indicate that the company is doing

NOTES

NOTES

well on the marketing front, it is not necessary to create hypothetical problems. The presentation must be completely from a neutral angle.

- Care should be taken that *only the relevant data is given*, which serves the purpose in fulfilling the objectives of the study.
- If a simple approach in analysis covering the more fundamental aspects serves the purpose, only to impress, *complicated analytical models should be avoided*. They confuse the sponsor more than being of any real help.
- If for some limitations, the sample size was small, *control the temptation of exaggerating*. Especially, in case of a large population, not only would it give wrong results, it could also be extremely misleading.
- *Give fair weight to all evidence*. This happens many times especially in industrial market research that the emphasis is only given on the feedback from large units while the small and medium scale units are overlooked. As it is quite likely that the consumption pattern could vary considerably in these two cases, both should be given due importance.
- While doing the screening of the data before presentation, *distinguish between opinion and facts*.
- *Continuity and completeness* of the data is a must
- It should be seen that the *data to be presented is reliable* and there is a secondary or primary source to which it can be contributed

12.3.2 Structure and Components of the Report

Having decided on the type of report, the report writer should now concern himself with its preparation. This can best be done when he is clear about what aspects or points are to be covered by it. Let us first consider the format of the research report.

Research Report Format

A research report can be written in a number of ways. However, three formats are generally followed. First, the report may use a *logical pattern*, which implies that the findings are presented in inductive order, i.e., moving from specific to general. Second, the report may follow a *psychological pattern*, which is almost inverse of the preceding pattern. In this format, the most critical information, i.e., the conclusion is provided first, after which follow the findings supporting the conclusion. Third, the report may use a *chronological format* wherein information is given along with the time dimension, i.e., things that happened earlier precede those which happened later. This form is generally combined with other formats. The chronological format is the least popular, though in respect of problems of a historical nature, is the most appropriate.

Having decided the format to be used in a research report, the researcher has now to prepare the report outline in accordance with that format.

Report Outline

Before attempting to write any report, the researcher must prepare a report outline. Without any outline, his report is bound to be haphazard and clumsy. An outline is to the writer what the *blue print* is to the construction engineer or what the *pattern* is to the dress maker. The outline will not only guide the writer to the order of the presentation of ideas but will also enable him to think before writing.

While preparing an outline, it must be noted that it should not be restrictive and rigid. It should be flexible so that if subsequently a change is needed, it can be easily introduced. Too frequent changes should be avoided. If they seem to be necessary, they will only indicate that the task of outlining the report was not done properly. It should also be noted that a change of one point in the outline may involve one or more related changes elsewhere for that point may have a bearing on one or more other points contained in the outline.

Changing one point may mean that other points are thrown out of sequence or out of context, so consider the relationship of a changed point to the rest of the piece and, where necessary, make changes in the other parts of the outline. The writer should, therefore, satisfy himself that all related changes as a sequel to one particular change have been carried out in the outline.

A typical report outline (contents) based on the logical format is presented as follows:

1. Title page
2. Letters of transmittal and authorization
3. Table of contents, statistical tables, charts and illustrations
4. Introduction
5. Review of literature
6. Methodology
7. Limitations
8. Analysis of data
9. Findings
10. Recommendations
11. Summary and conclusions
12. Appendices
13. Bibliography
14. Index

12.3.3 Style, Layout of a Report and Finalizing the Report

Any research report whether of a master's dissertation or thesis for doctoral degree or research project has to be presented in a scientific way with a standard design. The report should carry a title and date. It should contain a clear statement of objectives of research and an explanation to the methods followed. The text of report should be presented in a logical sequence and broken into readily identifiable sections. The findings should be summarized followed by suggestions. The layout of a report is as follows:

NOTES

NOTES

1. Preliminary text reference
 - (a) Title page (b) Approval sheet - Statement of the Problem (c) Preface - Significance (d) Acknowledgement - Need for the study (e) Contents - Objectives of the study (f) List of Tables - Hypothesis of the study (g) List of Figures - Definition of Concepts (h) List of Appendices
2. Review of literature
3. Methodology
 - Scope and area of the study - Aspects of the study
 - Period of the study
 - Methods of collecting data
4. Analysis of data - Tabulation
 - Figures
 - Tools used
5. Inferences
6. Brief description of the study - Findings
 - Conclusions and suggestions

The report should be presented in a readable and acceptable manner. The purpose of report writing is to suit the requirements of the clients for whom the report is being prepared.

Steps in Report Writing

The researcher follows the sequence of the content while preparing the research report. In writing a research report, there is no hard and fast rule, yet it should contain the information relating to the following:

- 1. Title page:** The title page refers to the subject of the report. It should be simple and brief. This page also indicates the name of the organizations for whom the report is being prepared. The date of preparation and the name of the persons writing the report are given on the title page.
- 2. Letter of transmittal and authorization:** This indicates the official who has authorized or sponsored the research and the official to whom the report is addressed.
- 3. Table of contents:** The table of contents is an outline of the order of the chapters, sections and sub-sections with their respective pages. If the report includes a number of charts, figures, tables, maps, diagrams and graphs, etc., a separate table for each category would immediately follow the table of contents.
- 4. Executive synopsis:** The synopsis or the summary is the most important part of the report because many executives will read only this part. It is a guide to those questions about which the executives would desire to know

more. It enables the executive to grasp the importance of the research. The synopsis precedes the details of the research report because it helps executives to read the summary first, and then read in detail if they require clarification on any point. The summary is a concise form of all the essential parts of the research report. It should include all the major facts, findings and conclusions. It involves background information containing important results and recommendations. The recommendations may not be in details but should contain all necessary information in broader headings and sub-headings.

The research report should begin with the problem undertaken for investigation. The problem is to be stated in a clear and logical manner enlisting the need, the values and significance attached so that the reader able to get a sufficient insight into the problem of study. The rationale or importance of the research problem for advancement of knowledge clearly as well as cleanly and the evils of marketing or constraints in the existing marketing theory are to be explained.

In brief, the investigation undertaken is to be spelt-out specifically and clearly. For example, if the research is undertaken on rural marketing it needs to specify that rural marketing, is a tradition-value system imbibed with social, economical, democratic features which have been experiencing many evils. Hence, there is a need for finding solutions and thereon framing a broad-based policy for effective working of rural markets. Investigation on this urgent issue is of highly essential.

5. **Introduction:** The introduction deals with the genesis of the report. It explains why the study was undertaken, the statement and formulation of the problem, formulation of hypotheses, if any, and the scope of the study. It is desirable to give a review of the related studies done earlier so that the reader may take the present report in proper perspective.
6. **Review of literature:** The importance of relevant literature and related previous studies are to be reviewed. This provides the background of earlier related studies and their investigation under different conditions, for the development of the present study and brings the reader up to date. A brief resume with integration of previous related studies provides a basis for identifying the gaps in the existing knowledge, (i.e., gaps between practice and theory, is to be included in this section. The researcher finds an opportunity to justify his own endeavour and to emphasize on the worthwhile elements of the study. A logical review of the literature brings the investigation results and points out areas or gaps that existed giving rise to the present investigation.
7. **Methodology:** The methodology adopted in the study becomes a significant part, on which the drawn conclusions are to be verified. The various aspects of the investigating problem say, the terms used for the purpose, scope and

NOTES

NOTES

area, research design, sample units, method and mode of collecting data, field work, tools of analysis, etc., are to be specified in clear terms with a logical approach. The limitation, if any, of the study is to be mentioned so that the nature of validity and reliability appreciated.

- 8. Analysis of data:** This section forms the crux of the research report. The collected data explained in previous actions are analysed and the results obtained are discussed. Data has to be classified, tabulated and studied using the scientific statistical techniques to get the results. The discussion should be critical, creative and revealing instead of being a transition of figures into words. A striking relationship between their variables must be brought out emphatically. The explanation should be explicit and supported by a pictorial presentation. The measure of presentation and analysis is a complicated task. The investigator has to use his patience, unbiased attitude, foresight and intellectual sharpness for this purpose.
- 9. Findings:** The findings of the study are presented in a logical sequence supported by adequate tables and findings. The findings may be presented in several chapters, sections and sub-sections. Each problem is discussed in a separate chapter. The irrelevant problems are ignored and relevant problems are thoroughly analyzed with the help of the collected information and data. The systematic presentation of the problems and their solutions based on relevant statistics have been a common form of the findings. Region, state, social strata, economic levels and other sub classifications have been the basis of presentation of the findings. Suitable tables, diagrams and charts are used to substantiate the findings. The details and basic statistics are given in the appendices. The findings are presented keeping in mind the objectives of the study.
- 10. Conclusions and recommendations:** The conclusions of the research-process are drawn by the researchers applying either the inductive or deductive method. The conclusions are scientifically tested statements and placed in a systematic manner. They verify the hypotheses or premises on which the research has been based. The conclusions are drawn according to the objectives and problems specified by the research. The step by step development of the conclusions is done under this section. The conclusions follow the natural findings. Conclusions denote that the researcher has complete knowledge of the research findings. Recommendations follow the conclusions. The knowledge of the situation and problem-finding solutions help appropriate recommendations. The researcher may recommend action only where it is necessary. Too much reliance on the recommendations is not permitted because the researchers would have made recommendations on the basis of specific situations. Therefore, wherever recommendations are made, the situations and limits regarding the same should be mentioned so that the executives may be aware of the situations and conditions for the implementation of the recommendation.

11. **Appendix:** The appendix is placed at the end of the research report. These items are the main stream of the report. The detailed statement may relate to the comprehensive data, formula, detailed information, such as the questionnaire, written interview instructions, etc.
12. **Bibliography:** The bibliography contains detailed information on books, references, journals and other materials relevant or referred in the course of the research work. The proceedings of the committees and references to the relevant portion of the findings may be given in this section. This is the last section of the report.

NOTES

Style of Report Writing

Research must include everything relevant and exclude everything irrelevant. The presentation must be clear and accurate with simple, short and coherent sentences. This requires skill as well as competence. A research report must not be loaded with phrases, idioms and long sentences. A researcher must describe his research findings in a scientific and logical style. An interesting report may appear dull if set in a complicated way while a dull report may attract interest if broken up into suitable sections.

The research report may be made more interesting by the use of such devices as frequent paragraphing, topical or summary sentences at the beginning or at the end of each paragraph, emphasizing the key phrases or sentences, use of specific examples, etc. The pronouns such as 'I' should be excluded and substituted by appropriate words. Vague expressions such as 'it seems', 'there may be' should be avoided. The past tense should be used in describing the research procedures that have been completed. The inclusion of graphs, charts and illustrations in the text of report should be seriously considered if it enhances presentation more clearly or forcibly.

A research report is a statement of findings summary from a research conducted with known scientific methods for a specific purpose. As such, it is different from a literary piece of work like an essay, a short story or a novel. Linguistic clearly is more essential for an accurate communication, but objectivity cannot be sacrificed for personal prejudices of the researcher.

12.3.4 Quality Research Report

The preparation of the research report is guided by objectives and methods of communication from the researcher to the clients. The report is tailor-made for the readers or clients. According to the nature of the reader, the report may involve technical sophistication, general interest in the area and many other relevant points. Reports are composed of words, numbers, graphs, charts and pictures. The researcher must be well versed in putting each of these components in such a way that the readers may easily grasp and understand them. If the researcher can understand the modes and interests of the clients, he can prepare a good report

NOTES

accepting or rejecting the findings; the researcher will start with the most interesting part of the research, avoiding expressions and tactics. The research makes colourful and vivid presentation of the report for instant decision-makers. The report will be very comprehensive with complete data for analytical types of executives. Dynamic executives require a concise and correct report. Besides these varied requirements, a plain and straightforward report must have some specific qualities. These qualities initiate certain *principles of report writing*.

1. **Easy to follow:** The body of the report should be written in a self-evident and easy form. Every subject should be presented under different heads. There may be different sub-topics and sub-headings under one heading. The statement may be presented in short and relevant paragraphs.
2. **Clarity:** Clarity in writing is as important as the quality of the communication. It is an art, which is acquired after considerable experience. The vagueness should be corrected because it may involve wrong decisions and faulty conclusions. The clarity may be observed by using a chart and marking sentences as well as avoiding too much elaboration. Difficult words are best avoided. Incorrect grammar may discredit a worthwhile report. A uniform style and format may increase the clarity of the report. When the underlying logic is imprecise and fussy, the readers experience difficulty in understanding the report and there may occur a possibility of confusion and misunderstanding. The original drafts are rewritten and finally arranged under specific devices. Use of too much jargon, ambiguous words and critical expressions may distort the clarity of the report.
3. **Conciseness:** The research report should be comprehensive as well as concise. The report should be long enough to cover all the components and objectives of the research project. However, this does not mean that it should be too detailed and wide. This should be avoided unless the report is required to be extensive. The researcher should not try to impress the reader with all the information at his disposal. He should try to eliminate all unnecessary information and also such information which is commonly known to people. Conciseness can also be obtained by the style of writing. Duplication of words, use of unnecessary phrases, poor expression and long sentences should be avoided. The researcher should express his findings in as few words as possible.
4. **Objectivity:** The objectivity of the report should be kept in mind throughout the report. Irrelevant reports are damaging and clients cannot fulfill their targets. The research objective is determined keeping in view the management's judgement. It has been observed that true research results have been avoided because they do not support the report. However, this does not mean that the researcher should include only those findings with substantial evidence and support. He should defend his results. The researcher may be asked to participate in the formulation of problems. He may frame

the objectives of the research accordingly and suggest to the management the solution of the problems. In many cases, the researchers are asked only to supply information rather than participating in the management. In such cases, they have to supply only information.

5. **Stress on practical report:** The report should be based on practical findings. It should not be a theoretical statement, the evidence supporting the findings should be given in the report. The researcher should have a realistic viewpoint. Writings should be restricted to practical findings and conclusions. Researchers should not write all those things, which are supported by the practical findings.
6. **Accuracy:** The accuracy of the report requires accurate information and data. The accurate information and data if analyzed properly and presented in a logical order will yield an accurate report. Accuracy needs careful handling of data, use of correct grammar, punctuation, spelling, etc.
7. **Completeness:** A report will be considered complete when it includes all information, which is required for the purpose of research. An incomplete report requires supplementary reports. If the report is formidable in size, it may not be useful. The interest and abilities of the researcher determine the completeness of the report.

Check Your Progress

1. State purpose which interpretation of data serves?
2. What is the report of research?
3. State three functions of a market research report.
4. List two aspects to be paid attention to before the presentation of the market research data.
5. What are the various aspects of the investigating problem?
6. What does the quality of accuracy of a report imply?

12.4 RESPONSIBILITIES OF A MARKET RESEARCH REPORT WRITER

The scope of marketing has expanded to include some very professional efforts by those who happen to sell ideas, behaviour patterns and other 'unconventional' products. Many of these products have strong social implications by themselves, and the involvement of marketing research in their success or failure is a matter deserving serious considerations. Marketing research can have a significant impact on society in ways other than the simple measurement of what the customer wants. In recent years, every election has had a well-oiled campaign in which marketing

NOTES

NOTES

research plays a key role. Marketing research is also increasingly applied during the legislative process, in formulating and evaluating legislation, and in judicial proceedings.

Accompanying the growing concern with marketing research in the political and legal settings, there remains the more traditional issue of ethics within the research process itself. Most responsibilities for ethical behaviour fall upon the researcher and the research user, since the primary duty of the respondent is simply to be honest in his behaviour and responses. Competitors and society at large are also indirectly involved, because their interests can be affected by marketing research activities.

Ethics is not simply a matter of compliance with laws and regulations. Ethics is more proactive than law. It is concerned with the development of moral standards by which situations can be judged. It focuses on those situations in which there can be actual or potential harm of some kind (for example, economic, physical or mental) to an individual or group. It is concerned with the development of moral standards that can be applied to those situations. It focuses on issues such as the following:

- Is the action or anticipated action arbitrary or capricious?
- Does it unfairly single out an individual or group?
- Does the action or anticipated action violate the moral or legal rights of any individual or group?
- Does the action or anticipated action conform to accepted moral standards?
- Are there alternative courses that are less likely to cause actual or potential harm?

A market researcher's most difficult ethical problems in dealing with the clients include the problems of maintaining confidentiality, maintaining technical and administrative integrity and ensuring that the research is applied correctly.

Ethical Issues

Marketing research involves an intellectual process of gathering and organizing information relevant to the marketing problem. Some respondents do not want to give information or they may be unable to provide such information. Is it ethical to get information in such situations? Should opinion be elicited from an individual who does not want to give his opinion? What measures should the research agency take to safeguard the respondents' confidentiality? What protection should the public have against invasion of secrecy or privacy on the part of interviewers? Is it permissible to mislead or deceive the respondents? Should a research agency become the respondent's partisan or advocate? In a way, ethics in marketing research covers a wide range of issues: some from the public as well as methods of the research industry, some with clients.

The public confidence, undoubtedly, may be improved if an ethical approach is followed to the marketing problem in its investigation. Marketing information, on which an industry is built and its fortunes are primarily dependent on the public. A company without marketing research ethics may make a 'fast buck' but it will land in a vacuum; from which it will not be possible to go further and progress.

NOTES

12.5 PRESENTING THE REPORT

Once the final draft of the research report is prepared and documented, the last stage is sharing the findings and research implications with the client or interested audience. This is usually done orally and with the support of visual aids. The presentation that the researcher might be making could be detailed for his team members or for an academic audience. However, in case the presentation is for the client or for a business audience, brevity and focus of the presentation is critical. A thumb rule for this is not to go beyond 20 minutes with more time for question and answers and interactive discussion on the findings.

Regardless of the audience for the presentation, the most critical aspect of the presentation is twofold:

- (a) Who is the listener? What does he/she seek from the presentation?
- (b) What is the core of the briefing—is it background, or methodology, key findings or decision directions that the findings are indicating?

Once the researcher is clear on this, he needs to need to focus on three key aspects:

- **Study background:** This should be essentially 10-15 per cent of the entire presentation. It should explain the impetus behind the study as briefly and with suitable emphasis as possible.
- **Study findings:** The major conclusions of the study need to be shared in simple words and with appropriate supportive visuals or material. The researcher must be able to demonstrate clearly the link between the study objectives and the findings.
- **Study implications:** In case this was agreed upon between the researcher and the client or was specified as a study objective by the researcher, this section would be the last section of the presentation. The link between what was found and what is suggested must be clear to the audience. The researcher may vary the discussion time between the earlier section and this as 45 per cent each or 30–70 or 70–30, depending on the study objective, i.e., more findings or more implication oriented. As supportive material the researcher can make use of:
- **Handouts:** These could be in the form of the primary questionnaire designed for the study or company brochures and other related secondary material.

NOTES

They should be distributed to the audience when the presenter is referring to them.

- **Slides:** These are created today with the help of computer programmes. There are endless possibilities enhancing the material to be presented and for engaging the listener. The designing and creation of the material requires considerable skill and care to ensure that the presentation style should be the supportive aid for an effective delivery and not a showcase of the computer graphics that the researcher is well versed with. Too much clutter and a random mix of text and graphics should be avoided. Animation of the data in synchronization with the vocal delivery makes the presentation more forceful.
- **Chalkboards and flipcharts:** These are additional visual aids that could be kept as standby for the question-and-answer session when an idea might have to be highlighted or demonstrated in the response of some query raised by the listeners. However, use of these means during an active presentation should be avoided as they necessitate the presenter to be engaged with the medium at the cost of losing contact with the listener.
- **Video and audio tapes:** Again, these are supportive materials that can be used to emphasize a point. The world has become smaller as a consequence of technological innovations that make dissemination of knowledge seem like child's play. Thus, the significance of communication and presentation of this learning cannot be overemphasized.

Check Your Progress

7. What is the last stage of report writing?
8. How should the study background of a report be?
9. What are chalkboards and flipcharts?

12.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Interpretation serves a dual purpose. First, it gives an understanding of the general factors that seem to explain what has been observed and second, it provides theoretical conception which can serve as a guide for further research.
2. The report of research is a statement of the contents in brief, the procedure adopted and findings arrived at by the researcher.
3. A market research report performs the following functions:
 - To communicate the marketing research result
 - To help the executives in selection of right action
 - To verify the conclusions of other reports

4. Before the presentation of the market research data, every researcher should pay heed to the following aspects:
 - While doing the screening of the data before presentation, *distinguish between opinion and facts*.
 - *Continuity and completeness* of the data is a must
5. The various aspects of the investigating problem are the terms used for the purpose, scope and area, research design, sample units, method and mode of collecting data, field work, tools of analysis, etc.
6. The accuracy of the report requires accurate information and data. The accurate information and data if analysed properly and presented in a logical order will yield an accurate report. Accuracy needs careful handling of data, use of correct grammar, punctuation, spelling, etc.
7. Once the final draft of the research report is prepared and documented, the last stage is sharing the findings and research implications with the client or interested audience.
8. The study background of a report should be essentially 10-15 per cent of the entire presentation. It should explain the impetus behind the study as briefly and with suitable emphasis as possible.
9. Chalkboards and flipcharts are additional visual aids that could be kept as standby for the question-and-answer session when an idea might have to be highlighted or demonstrated in the response of some query raised by the listeners.

NOTES

12.7 SUMMARY

- Data interpretation refers to the identification of trends in different variables. The researcher uses statistics for this purpose. The researcher is required to be familiar with the knowledge of the scales of measurement.
- The nominal measurement assigns a numeral value to a specific characteristic. It is the fundamental form of measurement. The nominal measurement calculates the lowest level of data available for measurement.
- The effectiveness of the research depends upon the methods of communication and presentation of the research report.
- Research is primarily a cooperative venture. It is essential that every researcher should know what others have found in their studies. Thus, it involves dissipation of knowledge and also creates ground for hypothesis and leads to further research on the same or an allied problem.

NOTES

- During the course of research, the researchers may prepare research papers suitable for publication in magazines, journals and newsletters. They may provide an insight into a problem.
- Reports are of three types. They are as follows:
 - (i) Report to the layman
 - (ii) Report to the administrator
 - (iii) Report to the technical expert
- *The technical report* is meant for the fellow professional / researcher, but not for the layman. This type of report should naturally contain all the details of the research study.
- If a simple approach in analysis covering the more fundamental aspects serves the purpose, only to impress, *complicated analytical models should be avoided*.
- A report may use a *logical pattern*, which implies that the findings are presented in inductive order, i.e., moving from specific to general.
- While preparing an outline, it must be noted that it should not be restrictive and rigid. It should be flexible so that if subsequently a change is needed, it can be easily introduced. Too frequent changes should be avoided.
- Letter of transmittal and authorization indicates the official who has authorized or sponsored the research and the official to whom the report is addressed.
- The summary is a concise form of all the essential parts of the research report. It should include all the major facts, findings and conclusions. It involves background information containing important results and recommendations.
- The findings of the study are presented in a logical sequence supported by adequate tables and findings. The findings may be presented in several chapters, sections and sub-sections.
- A research report is a statement of findings summary from a research conducted with known scientific methods for a specific purpose. As such, it is different from a literary piece of work like an essay, a short story or a novel.
- According to the nature of the reader, the report may involve technical sophistication, general interest in the area and many other relevant points.
- The body of the report should be written in a self-evident and easy form. Every subject should be presented under different heads.
- The report should be based on practical findings. It should not be a theoretical statement, the evidence supporting the findings should be given in the report.

- The scope of marketing has expanded to include some very professional efforts by those who happen to sell ideas, behaviour patterns and other ‘unconventional’ products.
- Ethics is concerned with the development of moral standards by which situations can be judged. It focusses on those situations in which there can be actual or potential harm of some kind (for example, economic, physical or mental) to an individual or group.
- The major conclusions of the study need to be shared in simple words and with appropriate supportive visuals or material. The researcher must be able to demonstrate clearly the link between the study objectives and the findings.

NOTES

12.8 KEY WORDS

- **Data Interpretation:** It refers to the implementation of processes through which data is reviewed for the purpose of arriving at an informed conclusion. It assigns a meaning to the information analyzed and determines its signification and implications.
- **Basic Report:** It is the first report prepared for the research project. It is prepared by the researcher for his own use as a working paper for further study and findings.

12.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is the purpose of a report?
2. How should a report for an administrator be?
3. What does the introduction of a report deal with?
4. Write a short note on review of literature.
5. What does the detailed statement of an appendix relate to?
6. How are ethics more proactive than law?
7. State the twofold critical aspect of report presentation.

Long-Answer Questions

1. Explain the classification of reports.
2. Analyse the three formats followed for making a research report.

3. Describe the layout of a report.
4. Discuss five principles of report writing.

NOTES

12.10 FURTHER READINGS

Saravanavel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.

Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.

Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.

Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.

Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

UNIT 13 APPLICATION OF MARKETING RESEARCH

*Application of Marketing
Research*

NOTES

Structure

- 13.0 Introduction
- 13.1 Objectives
- 13.2 Consumer Market Research
 - 13.2.1 Business to Business Market Research
- 13.3 Product Research
- 13.4 Pricing Research
- 13.5 Motivational Research
 - 13.5.1 Distribution Research
- 13.6 Advertising Research
- 13.7 Media Research
- 13.8 Sales Analysis and Forecasting
 - 13.8.1 Forecasting Methods
 - 13.8.2 Dating Mining
- 13.9 Answers to Check Your Progress Questions
- 13.10 Summary
- 13.11 Key Words
- 13.12 Self Assessment Questions and Exercises
- 13.13 Further Readings

13.0 INTRODUCTION

Marketing Research has a wide range of applications in business. It aims to analyze the demands and preferences of customers in addition to the current market scenario. Marketing Research does not include just one field of research, it has a number of areas which are assessed for acquiring an all-round knowledge of the prevailing circumstances and trends. These include product research, advertising research, media research and sales forecasting and analysis. In this unit, the aforementioned concepts have been explained in detail.

13.1 OBJECTIVES

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

- Discuss the concept of Consumer Market Research
- Analyse Product Research and Pricing Research
- Explain the dimensions of Motivational Research, Advertising Research and Media Research
- Describe the concepts of Sales Analysis and Forecasting

13.2 CONSUMER MARKET RESEARCH

NOTES

Many strategies are available in the process of researching consumers. For example, studies may differ according to the goal of the research, the type of data used, and the time frame of the investigation. Strategies of consumer research are classified according to their goals. The two major strategies of consumer behaviour are exploratory and conclusive studies.

(i) Exploratory Research

It is used to identify variables influencing consumers and discover how consumers may tend to react to these factors. This research occurs in situations when there is not enough data regarding consumers to draw conclusions about what variables are influencing their behaviour. Two significant methods used in exploratory research are consumer suggestions and focus groups.

(a) Consumer Suggestions

In the business world, many influences and problems encountered by consumers are discovered through the spontaneous suggestions of consumers themselves. For example, informal type of research similar to the familiar 'suggestion box' is conducted by many marketers. In this way, customers with complaints or new ideas are able to express them in a timely fashion.

(b) Focus Groups

Focus group interview is another popular technique for exploratory research. Focus groups generally bring together eight to ten people with similar backgrounds in a casual setting to apply the principles of group dynamics and free association to a marketing problem. A moderator guides the discussion but allows consumers to interact with each other. The sessions, which last about two hours are videotaped too. Focus groups can be helpful in specific ways to:

- (a) generate hypotheses about consumers and market situations.
- (b) suggest fresh and revitalised ideas.
- (c) check an advertisement package, or product concept to determine if anything about it is confusing, misleading, or negative.
- (d) understand consumers language and motivations.
- (e) understand consumers lifestyle and personalities.
- (f) explore a new area as a prelude to a quantitative study.
- (g) do a postmortem on a failed product.

It is important to appreciate that the primary objective of exploratory research such as focus group interviewing is hypothesis formulation, that is, forming a conjectural statement about the relationship between two or more variables.

(ii) Conclusive Research

Conclusive research builds upon exploratory research. Specifically the major roles of conclusive research are to describe consumers behaviour and to offer explanations for its causes. In addition, the predication of consumers behaviour and methods of influencing it can be suggested by conclusive research.

Type of Data Used

Two basis sources of data can be used in consumer research – primary and secondary data. Primary data are those the researcher gathers firsthand for the specific problem being investigated. However, there is a vast amount of information about consumers which is already compiled and readily accessible to the researcher who knows how to find and use it. Such data that have been collected for a purpose other than the research project at hand are termed secondary data. Before gathering primary data, the research should search through secondary sources to determine, if any are applicable to the problem at hand.

Research Time Frame

In consumer research studies, primary data can be collected either at one time or over a period of time. These research designs are referred to as a cross-sectional and longitudinal, respectively and the two approaches have different purposes.

Cross-sectional design is used to study behaviour at any one point in time. For example, such a study may aim to determine the knowledge various consumers have about a particular brand right after an advertisement. Interest can also focus on how some aspect of consumer behaviour occurs or changes over time. For example, marketers know that overall sales, as well as brand shares can undergo striking variations across weekly time periods. These studies are best studied by using a longitudinal design, which involves data gathering and analysis over a period of time. One popular type of longitudinal study is the continuous consumer panel. Consumers who are deemed representative of a particular group are chosen for inclusion and this fixed sample can be repeatedly studied. By asking the panel members the same questions over a period of time, changes in their behaviour as well as reasons for these changes, can be determined. Panel members generally maintain a continuous record or diary of their consumption activities, such as shopping purchase use, and product/brand decisions, as well as demographic and attitudinal characteristics.

Measuring Consumer Characteristics

Consumer research may also be classified according to whether demographic, activity or cognitive information is sought.

• Demographic Measures

Demographics research is concerned with gathering vital statistics about consumers – such characteristics as their age, income, sex, occupation, location, race, marital

NOTES

NOTES

status, and education. Since these characteristics are easily quantifiable, they enable the marketer to describe accurately and specifically and to understand better certain consumer characteristics. Much of the demographic data on consumers is obtained through government sources, especially through the census department. Population statistics and other official information are easily obtained and can be useful, even vital, for small businesses that cannot afford expensive marketing research.

• Consumer Activity Measures

Often the researcher seeks to understand various aspects of consumers activities. For example, questions such as the following may be of interest: When do consumers buy this item? which stores do they choose? How do they step in these stores? How often do they shop or buy? How loyal are consumers to certain brands? Which information sources do they rely on in making a decision regarding which brand to purchase.

• Cognitive Measures

Consumer researchers who desire to know more about their market than just demographic characteristics or activity pattern, may attempt to collect cognitive information, that is, information about consumers, knowledge, attitude, motivations, perceptions and information processing. Merely observing consumers cannot fully explain why they behave as they do, and questioning often does not provide reliable answers because of consumers' inability or reluctance to reveal true feelings to an interviewer. Thus, researchers attempt to utilize other techniques, some of which we have already seen. Exploring intervening variables is potentially useful in explaining consumer behaviour.

13.2.1 Business to Business Market Research

B2B marketing research is the process of uncovering insights into the marketplace by surveying a representative sample of its participants. Participants might include existing customers, former customers, prospective buyers, lost prospects (buyers who chose to buy from another company), and influencers. And in a competitive employer market, research might even include current and prospective employees, as well.

Check Your Progress

1. List two ways in which focus groups are helpful.
2. What is demographics research concerned with?

13.3 PRODUCT RESEARCH

Product research is essentially concerned with satisfying the need of the consumer in the best possible way by giving him the most optimum product.

Product research would set out to position the product on its attributes in a manner that the desired benefits are perceived by the customer. Depending upon the stage of life cycle in which the product is, the task of product research could be:

- (i) Development of a new product altogether.
- (ii) Improvements in product features of an existing product - product line extension, brand extension by evaluating what a product stands for.
- (iii) Introduction of new functional features in the product.
- (iv) Test marketing a new product.
- (v) Revitalizing a declining product.

Scope of Product Research

Product research includes the study of the following dimensions:

- (a) Raw materials used in various proportions in the final mix.
- (b) Attributes of the product.
- (c) Saleable points of the product.
- (d) Product planning.
- (e) Product appraisal.
- (f) Product classification.
- (g) Product design.
- (h) Product development.
- (i) Product packing.
- (j) Product branding or trade marks.

As products are meant for the people and as people are more important than the products, modern marketing is mainly concerned with the products which are matched suitably with the prospects. Product analysis is more concerned with the perception of the consumers about the product, than with the product as a bundle of physical and chemical attributes. Product research includes the study of planned obsolescence of products. Products grow older and exit, and make room for new products.

Object of Product Analysis

A product analysis refers to the study of:

- (a) The raw materials used.
- (b) Components.
- (c) Features like shape, size, colour, weight, design, etc.
- (d) The merits of the product - its convenience, portability, style, comfort, durability, wholesomeness or purity.

NOTES

NOTES

(e) Packing.

(f) Branding.

The main components of product research are product analysis and product development. A product is analyzed for what it has (merits), what it is (features), and what it does (functional utility and benefits drawn). The aim is product improvement in terms of its design, features, merits, uses, packing etc.

Sources of Product Research

These are:

- (i) Colleagues in the sales organization.
- (ii) Books and periodicals.
- (iii) Factory or workshop or plant.
- (iv) Casual sources (magazines, radio, television, movies, plays, etc.).
- (v) Competitors (through sales literature and advertisements).
- (vi) Consumers.

13.4 PRICING RESEARCH

Price is the main element in the marketing mix that produces revenue through the sales. Achieving the long-run pricing objectives always depends on sound price policy as well as price strategy. If the prices are high, buyers of the product are few and if the prices are low, many buyers purchase. Exchange of goods or services in terms of money is termed as *price*. Price includes cost of production plus the objective (profit) of the company. Efficient marketing job easily hinges on pricing policy for which prior decision on market-positioning is a primary factor. Thus, pricing is a delicate issue of decision making.

Issues Before the Company to Pursue its Objects Through Pricing Policy

1. Is company survival more important than its profit? As long as price covers variable costs and fixed costs, it must stay in business. Survival is only a short-run objective; the company in the long-run can learn how to add value or face extinction.
2. Can company exercise a sound assessment on estimation of demand and the costs associated with alternative prices and choose the price that produces maximum current profit, cash flow or rate of return on investment? Here, it faces the problems: (a) knowledge of product demand and cost functions in reality are difficult to estimate, (b) company puts more emphasis on current financial performance rather than long-run performance, (c) company ignores the effects of other marketing-mix variables, competitors, reactions, legal restraints on price.

3. Does revenue maximization lead to long-run profit-maximization and market-share growth? Price will be set to maximise sales revenue on estimating the demand function.
4. The company desire to maximise unit sales believing that a higher sales volume leads to lower unit costs and higher long-run profit. Price will be fixed at the lowest assuming the market price is sensitive. This is called *market-penetration pricing*.
5. Small market leader are expected by the company.
6. Is it right to set high prices to '*skim the market*'? The marketing management needs to thoroughly investigate all these issues for taking a sound and effective decision for setting pricing objectives.

NOTES

Factors Influencing Pricing

Internal and external factors which influence pricing of a product are listed here.

- Cost and profitability
- Nature of the market and kind of buyers – individuals and industrial buyers
- Extent of competition
- Objectives of the firm
- Marketing mix
- Product differentiation
- Buyer / supplier behaviour
- Elasticity of demand for the product or service
- Sufficient return on investment
- Stabilisation of prices

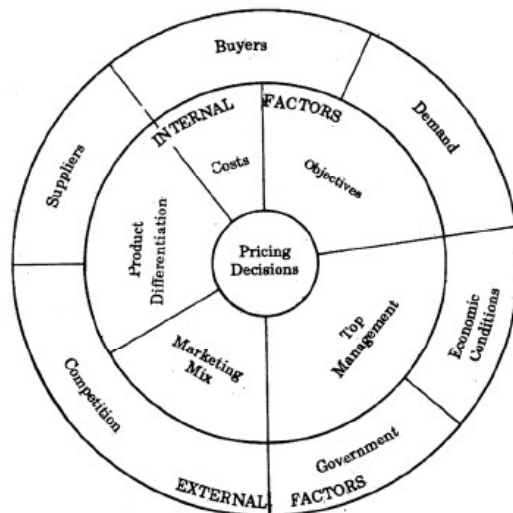


Fig. 13.1 Factors influencing pricing policy

- Channels of distribution used for distributing the product.
- Government control over maximum prices. In fact, there are many pricing policies in the light of which the final price for the product or service is fixed.

NOTES

Table 13.1 Forces that Push up and Pull Down Price

Forces that Push up Price	Forces that Pull Down Price
There is more demand but less supply.	There is more supply but less demand.
An increase in wages not matched by an increase in productivity.	Stable wages as productivity rises.
Weak competition and non-price competition.	Strong competition and competition on the basis of price.
Lack of efficiency in the use of land, labour, and capital by business.	The efficient use of land, labour and capital by business.
Buyers eager to acquire ownership (bullish attitude)	Buyer resistance of purchasing for any reason (bearish attitude).
Sellers hold goods for higher prices (bullish attitude).	Sellers eager to sell stock (bearish attitude).
The supply of money increases.	The supply of money decreases.
Goods are non-perishable by nature	Goods are perishable by nature.

Pricing Policies and Strategies

1. **Skimming the Cream Pricing:** A very high price is fixed for a new product at the initial stage. The main purpose is to make maximum profits from the initial demand. When TV sets were introduced a few years ago in India and recently, the Colour TV sets, the manufacturers fixed quite high prices to earn the maximum profits because the initial demand was very high. The purpose is to eat the cream before other competing manufacturers enter the line. The price is lowered at a later stage in order to maintain the demand for the product.
2. **Market Penetration Pricing:** This exactly is an opposite policy of the skimming the cream pricing strategy. The manufacturer tries to keep the price as low as possible in order to meet the competition to increase his share of the market for the product. The lower price naturally expands the demand for his product and increase the sales. Though the margin of profit is smaller the larger sales increase the profit of the manufacturer.

When to use a skimming or penetration price: Figure 13.2 summarizes when to use a penetration or skimming price. A skimming price strategy may be used when a new product is substantially different from competitive offerings and the market is not price sensitive and when the marketer wants quick returns.

A penetration pricing strategy may be used when the product is homogenous (not substantially different from other products), the market is price sensitive and the marketer first wants to get established in the market.

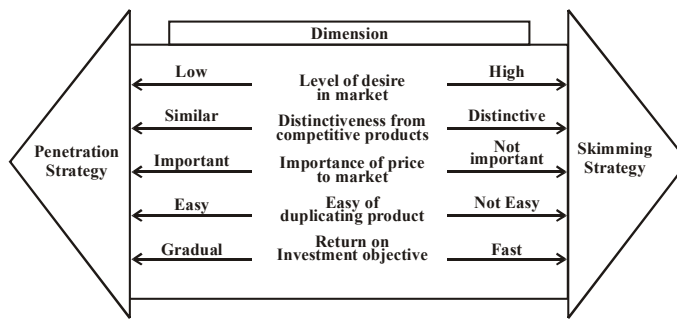


Fig. 13.2 When to Use Penetration and Skimming Strategy

NOTES

3. **Return on Investment Policy:** Any manufacturer who invests his funds in the manufacture of a product expects a fair rate of return on investment. Under this policy, the manufacturer first determines the rate of return on investment and then fixes the price accordingly.
4. **Customers Ability to Pay:** Ultimately the customers come to buy the product or service and pay for it, so the manufacturer has to consider the ability of the customers to pay, whether the price is high or low, the customers have to pay the money, hence the price must be acceptable to them. Many doctors, lawyers, chartered accountants etc., charge fees that their clients can pay.
5. **Extent of Competition:** The degree of competition prevailing in the market may be more or less, strong or weak. Similarly, the market conditions may be dominated by monopoly or competition or imperfect competition. Hence, after considering the existence and extent of competition, the manufacturer fixes the price first to survive and then to maintain the market demand for his product.
6. **Profit Maximization Policy:** In fact, every business activity is undertaken for the purpose of earning profit. Under this policy, the maximum price is charged to the customers which may lead even to the exploitation of customers. However, this policy can be followed only when the brand is well established and the competition is practically absent.

In short, the price fixation is a complicated affair and it does affect the sales of a product directly. Hence, continuous research is conducted to fix the price accurately for one's products.

Pricing Strategies for Initiating New Product

A company that plans to develop a new product problem regarding product position faces. It must decide where to position the product on quality and price. The marketer can follow any of the nine pricing strategies.

NOTES

		PRICE		
		High	Medium	Low
PRODUCT QUALITY	High	1. Premium strategy	2. High-value strategy	3. Superb-value strategy
	Medium	4. Overchanging strategy	5. Medium value strategy	6. Good value strategy
	Low	7. Rip off strategy	8. False Economy strategy	9. Economy strategy

Nine Product/Quality Strategies

Fig. 13.3 Pricing Strategies

Product Mix Pricing Strategies

Price is to be modified according to the product which is the part of the product mix. Here the firm wants to increase the profitability on the total product mix. Pricing becomes difficult when the various products have demand and cost interrelationships. There are six product mix pricing situations.

- (a) **Product Line Pricing:** Firms usually develop product lines rather than single product. While designing the price structure of different product items the firm should take into consideration the factors like (i) cost difference between (among) the items, (ii) customer evaluation of the different features and (iii) the competitor's prices. Here the marketer's task is to establish perceived quality differences that justify the price differences.
- (b) **Optional Feature Pricing:** Most firms offer optional/accessory products or features along with their main product. The pricing strategy is to keep the prices of the optional product on the higher side comparatively.
- (c) **Captive Product Pricing:** Companies that make products which must be used along with a main product, use captive pricing. Examples are razor-blades, camera-film, computer-floppy. The prices of main products are kept on the lower side and the prices are comparatively higher for the next part.
- (d) **Two Part Pricing:** In case of services, the captive product, pricing strategy is called as two part pricing. Service firms often charge a fixed charge plus or variable image usage fee e.g., telephone pricing, warehouse charges etc.
- (e) **By-Product Pricing:** In pricing of some products, there are some residue products which are called as by-products. If the by-products have little value and are in fact costly, the manufacturer should accept any price that covers more than the cost of disposing them. The revenue earned on the by-products will make it easier for the company to charge a lower price on the main product if the competition is intense. For example, we get petrol from the processing of crude oil by-products are diesel, kerosene oil, coal tar etc. The prices of products are kept on the lower side.

- (f) **Product-Building Pricing:** It calls for setting the prices of a bundle. Bundle pricing is less than the individual item pricing. Price bundling can promote the sales of the products. Consumers might not otherwise buy, but the combined price must be low enough to get them to buy the bundle.

Price Adjustment Strategies

Companies do not set a single price but rather a pricing structure that reflect variations in the changing situations. Price adjustment strategies include:

- (a) Geographical Pricing
- (b) Discount Pricing and Allowances
- (c) Discriminatory Pricing
- (d) Psychological Pricing
- (e) Promotional Pricing

Pricing Objectives

If the marketing management is to perform its operations efficiently, it has to set pricing objectives. The objectives of pricing must be included in the pricing policy. The main objectives of pricing which are logically related to the company's overall objectives are discussed hereunder.

- (i) **Business needs return on investment:** Businessmen expect (probable) return on their investment. Many well established companies follow the pricing objective in terms of return on investment (ROI). The pricing objective is also known as *pricing for profit*.
- (ii) **To maintain or improve Market Share:** In pricing the products, the potential 'market share' is an important consideration. A good market share is better indication of progress by lowering the price. Consumers are benefitted. To know whether the market share is increasing or decreasing, the management has to compare the present market share with the past market share.
- (iii) **To prevent competition:** While fixing the price, the price of similar products produced by other companies will have to be considered. One has to look to the price of rival products and then chalk-out proper price to meet market competition. There is a need to fix price at a lower level at the time of introduction of new products to the market and this low price attracts customers on a wider scale and establishes good market share. Further, this low price policy discourages the competitors, but should be adopted before the obsolescent phase under product life cycle.
- (iv) **To maximize profit:** Profit is the primary motto of business of all kinds. Maximization of profit should be on total output but not on one item. Fixing the prices to products during the period of depression should be such that the prices are not allowed to fall below a certain level; and in the boom

NOTES

NOTES

period, the prices are not allowed to rise beyond a certain level. Therefore, pricing is a long-run objective and aims at preventing frequent and violent fluctuations in price.

(v) **Mobilization of resources:** The products are priced in such a way that adequate resources are made available for the company's expansion and developmental investment. Marketers are keenly interested in getting back their investment as speedily as possible.

(vi) **Other objectives:** These include achievement of price stability, survival in the market, building public image etc.

Market Demand Survey

It is to examine the pricing objectives. No specific procedure is adequate to be made applicable to all companies to determine price. A thorough survey could be made to know demand and thereupon fixing the product price considering all legal obligations.

Check Your Progress

3. List some tasks of product research.
4. What is a product analyzed for?
5. What is market-penetration pricing?
6. When can a skimming price strategy be used?
7. What should a firm take into consideration while designing the price structure of different product items?

13.5 MOTIVATIONAL RESEARCH

Interest in consumer attitude measurement now seems to be at an all-time peak. Most of the strategy of market segmentation is based on attitudinal segmentation. Attitude measurement is the key factor in the increased efforts to measure the effectiveness of advertising. A knowledge of attitudes helps the marketer predict consumer reactions to products and to advertising messages.

Motivation Research "is a term used to describe the application of psychiatric and psychological techniques to obtain a better understanding of why people respond as they do with products, advertisements and various other marketing situations." It was after the Second World War that the motivation research specialists appeared on the business scene suggesting the use of psychological techniques in marketing research particularly in the field of ascertaining consumer attitudes, preferences, motives etc. Motivation research attempts to gather information buried in the unconscious mind of the consumer. The basic methods used by motivation researchers are Projective Techniques and Depth interview.

Nature of Motivation Research

*Application of Marketing
Research*

Motivational research, by its very nature, is different from traditional research. The research is directed at finding the answer to a specific question, or even the number that answers a specific question. The biggest problem in motivational research is that direct approaches generally do no work. The researcher cannot simply go up to a consumer and ask him why he prefers one product brand over another and expect to get a correct answer. Generally, direct questioning does not give correct answers for one of the two reasons : either the consumer does not know his true reasons or he knows but is unwilling to tell them to the researchers or gives illogical reasons. Let us explain these reasons in the following pages.

NOTES

1. **The Consumer Does no Know The “Why” of Purchase:** Consumers are generally unaware of their own “true” reasons for specific brand and product selections. For instance, cigarette smokers are often extremely brand loyal. If you ask them why they prefer specific brands, they will answer without hesitation, because of the “Taste”! Unfortunately, this answer does not square with the well-documented fact that the average consumer is unable to distinguish among various same-size cigarettes. This can be tried out by putting a cover on the brand names of different brands of the cigarettes. Now, these cigarettes with their brand names covered may be given to a person who boasts of having developed brand loyalty for cigarettes on the basis of taste at random, it will be extremely difficult from him to identify his own brand out of the ones supplied to him. So we can conclude that the smokers are brand loyal not to a specific taste but to a specific cigarette image or corporate image of its manufacturer. Thus, if we are interested to know why someone is brand loyal to a particular cigarette (or almost any similar product, viz., beer), we should not ask him directly. Instead of this, we must investigate his attitudes toward various brands. Thereby we can indirectly determine the images, and personality profiles of the consumers pertaining to concerned brands.

There exists another reason of consumers not knowing the true reasons for purchasing and subsequently their failure to give accurate answers to direct questioning. Regardless of the specific purchase situation, they tend to repeat the answers that have been conditioned by the advertising. The consumers can also get confused about the brand they actually bought. This happens because of the influence of heavy product identification (generic names) with a single brand. They may consider all facial creams as “Ponds”, all soft drinks as “Coke”, all toothpastes as “Colgate”, all photocopiers as “Xerox”, and so forth.

From the standpoint of the major manufacturers, this heavy brand product identification (identification through generic names) can be viewed either a blessing or a curse. This depends upon the objectives of the firm’s management. But from the standpoint of market researcher, it is a total

NOTES

disaster. In response to direct questioning concerning purchase behavior, all that the researcher gets is a major brand name-even when the consumer is buying “off-brands”.

2. *The Consumers Will Not Tell About Their Purchase:* Another reason for incorrect answers lies in the choice made by many consumers, not to tell the correct answer even if they know it. Generally speaking, it can be assumed that a significant amount of deliberate misrepresentation will occur at any time the subject area is either morally sensitive or status-related. Thus, people tend to overestimate their income, their charitable giving, and their purchases of name brand items, and they tend to underestimate their drinking, their smoking, their adultery, and their purchases of house-brand items. Respondents will also modify their answers in an attempt to appear logical and intelligent. For instance, the purchaser of a new “Bullet motor cycle” may know that the reason he bought such an expensive motorcycle was to show it off to his friends and relatives, but it is doubtful he will say so to the interviewer. Rather he will talk of the outstanding style, speed, economy, safety and good resale value of the mobike. His true reasons can be discerned only through the use of the relatively indirect methods of motivation research.

3. *The Consumers May Put Forward Illogical Reasons for their Purchase:* The reasons for consumer preferences are seldom logical. The researcher who approaches the respondent in a direct, logical, and straightforward manner will often obtain very misleading answers. Whereas, some motivation research can be conducted in a relatively direct fashion while in other areas, valid results can be obtained only from very indirect and seemingly roundabout methods. Furthermore, since the nature of investigation is subjective, the conclusions produced are often highly speculative and will carry greatly with the person performing the research.

Kinds of Information Sought in Motivation Research

We need to “....focus attention on the whole battery of inner conditions that play a dynamic role in a person’s buying or not buying, responding favourably or unfavourably to some communication. The battery of inner conditions includes “.....cognitive factors such as information or misinformation people have, and perceptive factors such as the way they see a situation. These factors as well as the motivational forces such as attitudes, expectations, habits, and intentions-all influence the way the consumer behaves.

We may not be able to pinpoint at this juncture what information is relevant in the study of motivation research, but any information which has something to do with the why part of the consumer behaviour is significant in the motivational research. All information related to perceptions, learning, ethnic relations, culture, society, personality, and group influences have to do something without motivation

research. Therefore, the information on the entire field of human behaviour is needed for conducting motivation research. It is difficult to say which information should or should not be included for carrying out such study. In the ensuing text of this section, we will focus our attention on some of the major fields of human behaviour regarding which information is commonly sought while conducting motivation research. These areas are, namely, attitudes, assumptions, sensations, images and motives.

13.5.1 Distribution Research

Distribution research refers to the collection and analysis of information related to the sales of a product or brand and its distribution through various retail channels so as to enable the management make better decisions. In depth distribution, research about a brand/product provides information related to its retail presence, market size and share, sales achieved, how well the competitor is selling and seasonality of demand.

Marketing and distribution research is a key means to understand and analyze the marketing environment and thereby helping eliminate any weak links present. Distribution research helps gain insights into the following:

- Better management of sales channels with enhanced knowledge
- Better segmentation of distribution within sales channels
- Roles played by intermediaries in the sales process
- Understanding the centers of influence within a sales channel
- The company's market position within a particular sales channel

Check Your Progress

8. Why does direct questioning not give correct answers?
9. Why is brand product identification a disaster from the standpoint of the market researcher?
10. What information does research, in depth distribution, provide about a brand?

13.6 ADVERTISING RESEARCH

The advertising research is an application of marketing research aimed at the measurement of advertising effectiveness (how well an advertisement or advertising campaign accomplishes its objectives) and ameliorate advertising efficiency (the most optimum use of the advertising budget and media-mix in implementing an advertising campaign). The primary aim of advertising is to sell an idea, a good or a service whereas the ultimate goal in advertising research is to measure advertising impact or influence on sales of that idea, good, or service.

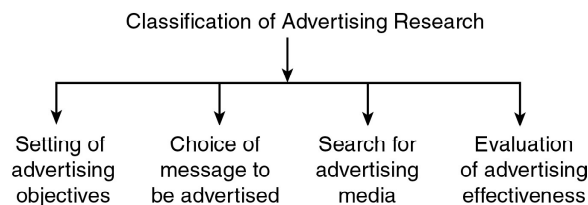
NOTES

The scope and coverage of advertising research include research in advertising objectives and product appeals, (message) copy testings, media selection, advertising effectiveness, and advertising budget / expenditure.

NOTES

Classification of Advertising Research

Advertising research can be classified in the following categories.



- (i) **Setting of Advertising Objectives:** This implies the specification of what people demand from advertising research. This deals with information about already developed advertising methods, improvements in advertising strategy, markets etc. The advertising objectives can be framed in terms of awareness, changing attitudes, changing predispositions to buy or some combination of the three.
- (ii) **Choice of Message to be Advertised:** This aspect of research conducts tests with regard to the total advertisement, its components and the dominant theme. Here a decision is made by asking some searching questions viz., what their message should be, how the message should be communicated to the consumer, where, when and to whom. Information regarding attitudes, awareness and perception of consumers regarding the product characteristics can be useful guide for providing effective message formulation.
- (iii) **Search for Advertising Media:** Reaching the right people at the least cost is the aim of every advertising agent. There are many types of advertising media viz. newspapers, radio, magazines, television, posters etc. Media research attempts to eliminate waste in advertising by objectively analysing the media available for promoting products and services. Media research is concerned with measuring the size and composition of individual media vehicle.
- (iv) **Evaluation of Advertising Effectiveness:** This research is concerned with analysing different media or their combinations and evaluating the degree of success with which the advertising objectives have been achieved.

Objectives of Advertising Research

The following are main objectives of an advertising research investigation.

- (a) How will an advertisement accomplishes its objectives?
- (b) Optimum utilisation of advertising budget and media-mix in implementing an advertising campaign.

- (c) Measurement of advertising impact on audience.
- (d) To bring cost effectiveness in advertising.

Setting of advertising objectives and budgets: Basically, the advertising must relate to consumer behaviour i.e., objective should be in terms of changing consumer attitudes with respect to selected product characteristics. The goal of advertising often centers on attempting to alter the attitudinal set for the brand in question within a given segment in the direction of those product characteristics which were considered to be salient. The study requires measurement of the attitudes of target audiences towards the product and then to evaluate the effect of advertising on these attitudes. Information is needed about the whole aspects of behaviour which surrounds the purchase and use of the product, role served by the product, the system which is employed by the user to obtain the desired goals, the events for which certain products are appropriate, the standards employed by the consumer whether the product is or is not performing satisfactorily and how the product is perceived in terms of functional equivalents. In addition to this the following information is also necessary:

- (a) Pre-advertising measurement of level of awareness, attitudes and predisposition to buy.
- (b) Measuring the relationship between the communication objectives, sales and profits.

The advertising budget will depend on the nature of tasks and the objectives formulated. It is the total of the costs associated with each task. The following are the main steps in formulation of advertising objectives and budget:

- (a) Setting up operational advertisement objectives.
- (b) Determining the tasks that have to be performed to attain these objectives.
- (c) Arriving at the advertising budget by totalling the costs of performing the tasks.

Product Appeal Research

Advertising research is concerned with measuring the attitudes of consumers towards whom the advertisement is targeted. After measuring the attitudes, an attempt is made to determine the effect of advertising on these attitudes. The consumers might have favourable or unfavourable attitudes towards competitors' brands in comparison with our brand of a particular product. The attitudes can be measured to certain degrees of accuracy and they are the predictors of behaviour of the individuals. On the basis of this data, the attitude of customers towards both the brand and the competitors' brands can be determined. Accordingly, a marketer can determine what attitudinal changes are required to be brought about in order to convert the unfavourable attitudes into favourable ones towards his brand.

It is believed that the average consumer has a reasonably stable set of attitude towards salient product characteristics pertaining to a product class. With this

NOTES

NOTES

stable set of attitudes called "*backdrop*", he evaluates one brand of the product as compared with other brands. There exists a difference in perceiving the salient characteristics of a product in question in different segments of a market. The researcher's task is to find out the salient features of a product as perceived by each segment of the market. Once these salient features are ascertained, product appeal is designed in order to change the unfavourable attitudes into favourable ones. Hence, the researcher ventures to change the unfavourable image of target audience (towards his brand of the product) into a favourable one through the modification of the product appeal. It is not essential to criticise the competitors' brands but to highlight the good qualities of one's own brand.

To illustrate, we can consider an example of toilet soaps. Suppose, in market A the salient characteristics of toilet soaps may be:

- lather generating capabilities
- perfumes or pleasant odour
- skin protection
- unaffected shape till the end.

But if first two of these characteristics are considered salient in market B, then the emphasis of the product appeal in markets should be on the first two characteristics only. Similarly, in other markets, some other sets of features may be treated as salient ones. So, stress is required to be laid on the salient features of a product in a particular market segment and the product appeal in advertisements must be designed accordingly. The aim is to reinforce the salient features amongst the present users. An attempt must be made to measure the attitudes by highlighting the salient features in product advertisements. Thus, for improving the product image, market segment's salient features must be pinpointed and the product appeal must rotate around these.

Advertising Message (Copy Testing) Research

This focuses on the ability of the advertisement to achieve impact and to project the desired message. Information regarding attitudes, awareness and perception of consumers regarding the product characteristics can be useful guide for providing effective message formulation.

Advertising message research needs measurement of consumers' attitude on whom the message is targeted and after that, determining the effect of advertising on these attitudes. There are some stable set of attitudes within each consumer towards some product class from which he evaluates one product brand with other brands. There exists a difference in perceiving the salient characteristics of a product in different segments of a market. The researcher should identify these salient features of each segment and then design the product appeal/message in order to change the unfavourable attitude into favourable one. It is not essential to criticize the competitor's brand but should highlight the good qualities of one's own brand i.e., in designing advertising messages, stress is to be laid on the salient

features of the product in a particular market segment. This requires more elaborate techniques like depth interviews and projective tests. Other tests may be devised to check the spontaneous emotional response to advertisements and include physiological measurements of different kinds viz., eye-blink rate, pupil dilation, breathing rate etc. Eye blink rate indicates the intensity of interest and is measured on special equipment.

Post-testing of advertisement develops on the basic measurements of recall recognition. Tests to measure recall (verbal or pictorial) may be either aided or unaided. Respondents in aided recall test may be shown a series of advertisements. After these have been viewed for a specified period, the respondents are questioned about the advertising message, product brands etc.

Messages may also be sharpened in order to support existing beliefs. Consumers may add subjective meanings to advertising messages which advertisers had not intended should be read into their advertisements.

Check Your Progress

11. What does media research attempt to do?
12. What does the goal of advertising center upon?
13. State the main steps in formulation of advertising objectives and budget.

13.7 MEDIA RESEARCH

Media research is an important component of marketing research. The media are vehicles which carry the advertisement message to the target market, the media being newspapers, magazines, radio, T.V., outdoor advertisements, etc. In each medium category, there are so many individual media. Media research attempts to evaluate and determine the most suitable and convenient media for the advertiser. It attempts to determine and assess the reach and impact of various media. There are however, other considerations, too, such as media selectivity, memory value and the cost of the media.

Proper media research is nowadays necessary for any advertiser who is spending a fortune to advertise his products. He is interested in knowing the demographic, economic, social, psychographic profiles of media audiences. He is interested in rating the media region-wise, in knowing the reach or circulation of each medium, in rating the editorial features of different media. Media research agencies do media research for the advertisers.

Audit Bureau of Circulation (ABC): This is an agency which gives certification to media circulation against a fee. ABC also functions in India. It collects its information from the vendors of the media. Besides, it conducts its own surveys.

Private market research agencies in media research: The Indian Market Research Bureau conducted the first National Readership Survey (NRS-1). Many

NOTES

NOTES

years later, MARG, another Bombay-based agency, conducted NRS-II. Then, NRS-III was jointly conducted by IMRB and MARG. Later on, NRS-IV was conducted independently by ORG and MARG. Doordarshan has its own market research department, which conducts viewership surveys. We need an NRS every year in our country to give a boost to advertising activity. IMRB conducts TRP studies for TV programmes.

Functions of Media Research

The primary function of media research is to select the media most efficient for our ad campaign (media mix research). It is a part of media planning. Here the research is concerned with what type of media should be used. Magazines? Newspapers? TV? Outdoor? If magazines, of what classes? General interest magazines, women's magazines, crime magazines, business magazines? If general interest magazines, which specific ones – India Today, Sunday, Onlooker, the Week? If newspapers, then in which regions? If radio or TV, which coverage is in mind-national network or regional? Once media selection is made based upon our research data regarding product, market, target audience, demographic and geographic and psychographic factors, the next logical step is to determine by research the combination of 'mix' of the media one must use to achieve the desired promotional objectives. Here, the media and the target market are to be matched. The overriding factor is that of the ad budget. A balanced mix can be achieved by the right type of research. It considers the research inputs of frequency and reach. Research guides us whether to use one type or varied media mix. Research also tells about the 'wear out' effect when the message is repeated and people become bored with it. We also know by research, the media impact, e.g., heavy watchers of TV are light readers of magazines, and vice versa. Concentration on TV advertising alone will have greater impact on heavy viewers and very little impact on light viewers. The impact is uneven. It is to be balanced. Media mix research also covers the duplication effect, gross audience concept and media scheduling. These days complex mathematical models (like LP and simulation) are used in media mix research.

There can be a problem of media measurement, the parameters of which are reach, frequency, yield per rupee and effectiveness. Out of this, effectiveness of media is very difficult to measure. The researchers are trying to develop a method to equal all parameters on different media.

Media's reach is expressed in terms of '*vehicle distribution*' which refers to the number of copies circulated within a region/country. In case of broadcast media, it refers to the number of programmes distributed on each channel.

'*Vehicle exposure*' refers not only to the number of people but also the type of people reading an average issue of the media. It also includes the frequency - the number of times each person reads an average issue.

'*Advertising exposure*' refers to the number of people exposed to the ad message and the frequency of exposure.

13.8 SALES ANALYSIS AND FORECASTING

Sales forecast is an estimate of sales in monetary or physical units for a specified future period under a proposed marketing strategy or programme or under an assumed set of economic and environmental forces, planning premises outside the business organization for which the forecast or estimate is made. It is based on an effective system and is valid only for the specified period.

Need for sales forecasting: The sales forecast should be one of the company's most important documents. It is the pivotal factor around which most business planning rotates. Important decisions like manpower requirements, advertising and other promotional outlays, etc., are dependent on sales forecast. Any significant forecasting error can have serious and far reaching consequences with regard to the performance of an organization.

Sales forecasting is a complex phenomenon and is based on a variety of concepts and techniques. In literal sense forecasting means prediction. It may be defined as a technique of translating past experience into prediction of things to come.

The following are the main *components of a sales forecasting system*:

- (i) Market Research Operations to get the relevant and reliable information about the trend in market.
- (ii) A data processing and analyzing system to estimate and evaluate the sales performance in various markets.
- (iii) Proper co-ordination of steps (i) and (ii) and then to place the findings before the top management for making final decisions.

The forecasting programme should not be wholly identified with a single personality. It should be organized by marketing research manager but must have the active participation of sales and production managers.

Due to dynamic nature of market phenomenon, sales forecasting has become a continuous process and requires regular monitoring of the situation. The overall success of an enterprise mainly depends on the quality and reliability of sales forecasts.

Sales forecasting is becoming more and more important function in modern times, as our abilities in this area are continuously improving and the organizations are leaning more towards planning. The increasing availability of economic data, the increase in computational ability and capacity to deal with more factors through computers have made it possible for organizations to forecast their sales with greater accuracy. The new statistical approaches are improving with great speed, essentially due to the computer. Consequently, sales forecasting is one of the more dynamic fields of business.

NOTES

NOTES

In this section some typical approaches to the problem of sales forecasting have been discussed.

Steps in forecasting: The following are the main steps in sales forecasting:

- (i) Determine the objective of the forecast
- (ii) Identify the period over which the forecast is to be made.
- (iii) Select the technique to be used for forecasting.
- (iv) Collect the information to be used.
- (v) Make the forecast.

Objectives of sales forecasting: The objectives of sales forecasting can be summarized as follows:

A. Short-term objectives

- (i) Formulation of production policies
- (ii) Availability of regular supply of raw material
- (iii) Full utilization of machinery
- (iv) Determination of price policy
- (v) Regular availability of labour
- (vi) Control of sales

B. Long-term objectives

- (i) Planning the plant capacity
- (ii) Provision of labour
- (iii) Long-term production planning
- (iv) Success of managerial planning

Importance of Sales Forecasting

Sales Forecasting is the most important aspects of marketing programme of company. It is not only the commencement of sales budgeting but of the budgetary control as a whole, In the words of J. Batty, "For many types of business the sales forecast is the commencement of budgeting. The quantity of the company's product which can be sold may be the principal budget factor, and therefore, around it revolves the whole planning process. When the demand is assured, selling is principally a matter of producing as much as possible at the right cost." Sales forecasts are the basis of controlling the sales promotion also. They are the main basis of setting sales volume quotas, since carefully prepared sales forecasts when intelligently broken down result in quotas that are reasonable and thus attainable. By breaking the sales forecast down into manageable parts, i.e. into sales volume quotas for individual salesman-management can specifically define the results it expects from each man's efforts. As a matter of fact, sales forecast is the basis of future management planning. A sales forecast involves more than an appraisal of

conditions in the product markets. It also reflects the manner in which other environmental factors, such as resources, markets and government policies may influence sales objectives and it reflects the limitations too that evolve from the productive capabilities of the organization. As a planning premise, it is regarded a key to internal business planning. It is only on the basis of sales forecast that business and capital outlays and various policies are formulated and finalized. Hence, sales forecast should be made very carefully.

Factors to be Considered in a Sales Forecast

In the preparation of a sales forecast, the sales people should take into consideration the following factors:

- (1) **Past Sales Figures and Trends:** The record of previous year's sale is the most reliable basis as to future sales as the past performance is based on actual business conditions. Hence past sales figures must be analyzed to determine the future trends, other factors affecting future sales, e.g., seasonal fluctuations, growth of market, trade cycles, etc., should also be considered.
- (2) **Salesmen's Estimate:** The salesmen's estimate should also be given due weightage while preparing sales estimate for the whole concern. It is fruitful to consider the estimates of sales received from salesmen because they can make more accurate estimates, being in direct contact with the consumers. However, it should be observed that salesmen's estimates should neither be overoptimistic nor too conservative.
- (3) **Plant Capacity:** The sales budget should be within the plant capacity available and should ensure proper utilization of plant facilities. Proposed plant extensions should be allowed for in the preparation of the sales budget.
- (4) **Availability of Raw Materials and Other Supplies:** Adequate supply of raw material and other supplies should be ensured before preparing the sales estimates. Sales estimates should be adjusted according to the availability of raw materials if the raw materials are in short supply.
- (5) **General Trade Prospects:** The probability of the sales going up or down depends on the general trade prospects. In this connection, valuation information may be gathered from financial newspapers and magazines such as *Economic Times*, *The Financial Times*, *The Commerce*.
- (6) **Orders on Hand:** In boom periods or where production is a very lengthy process, the value of orders on hand may have considerable influence on the amount of sales to be budgeted.
- (7) **Seasonal Fluctuations:** In the preparation of the sales budget, seasonal fluctuations should be considered because sales are affected by these fluctuations. In order to have an even flow of production, efforts should be made to minimize the effect of seasonal fluctuations on sales by giving special concessions or added inducements during the off-season.

NOTES

NOTES

(8) **Financial Aspect:** The sales budget should be within the financial capacity of the concern. Sales expansion usually requires the increase in capital outlay. Thus, if any big sales expansion is planned, it must be ensured that facilities are available to finance the operations.

(9) **Adequate Return on Capital Employed:** The sales volume budget should produce an adequate return on the capital employed.

(10) **Competition:** To have a realistic sales budget capable of being achieved in the face of competition, the nature and degree of competition.

13.8.1 Forecasting Methods

The sales forecast preparation requires the availability of the sales turnover of the company for the past years, so that the company can understand the sales patterns. The company should also have a clear understanding of all the factors that are likely to have an impact of the sales in the forthcoming year.

Mathematically, it is possible to forecast sales with some precision. Realistically, however, this precision can be dulled because of external market and economic factors that are beyond your control.

The main advantage of the sales forecast technique is that it helps the company to plan its material and resource requirement for the coming year based on the anticipated sales volume. Therefore, the company can alter its production schedule, capacity, marketing resource requirements, personnel requirements, etc., on the basis of the forecasted sales. For instance, if the circumstances indicate that the sales for the first six months would be lower than the previous year, the company can cut down its production and lower capacity utilization, instead of piling up inventory. This usually happens during the period of recession. During the same period, sales and marketing efforts may have to be higher to achieve the forecasted sales volume. Accordingly, resources would have to be reallocated within the company. The company can also then look at revamping its pricing strategies for the same period. It can increase its promotional offers, and give additional incentives to its channel partners to push its product to the customer.

Forecasting helps the company to allocate sales quotas per region, territory, branch and salesperson. Further, it also allocates quotas for every customer segment, or individual customer (for large customer accounts). This helps the company to plan its distribution and communication budgets. The company can plan its entire marketing mix on the basis of the sales forecast. The product mix, product features, pricing, promotion, and distribution decisions will be made on the basis of the sales forecast. Resource planning is also done by using the sales forecast. Any resource requirement can be planned at this stage to ensure smooth operations for the entire year.

For the forecasting to be accurate, managers need to consider all of the following factors:

*Application of Marketing
Research*

- Past sales trend: The first data that the company needs to have is the sales turnover for the last many years broken down into marketwise, productwise and customer segmentwise. These data will form the most fundamental basis of any sales forecast. The sales turnover figures need to be available for a considerably long period, around a decade. The reason is that a longer time period without aberrations is caused by temporal factors. Therefore, the future predictions are likely to be more accurate. In case a company's life is shorter than the minimum requisite time period, it can take an approximation from an almost similar firm in the same industry, though results will not be very accurate (since no two companies will be exactly similar). A longer time period also allows the company to discern patterns in sales variations, that can be traced back to a few causative factors.
- A company needs to understand clearly that its ability to translate the forecast into actual sales depends on the level of critical resources (such as production, marketing, finance, personnel) available with it. For instance, an optimistic forecast is useless, unless, the company can back it up with a sound plan with optimal resource allocation.
- The company's market share, public image and position vis-à-vis competitors also play a very important role in realizing its sales forecast. The company must know if it has a loyal customer base, and it must possess data that reveals the profitability of its customer base.
- The economic environment of the business is undoubtedly one of the most important factors that affects the sales turnover of a company. The forecast must take into consideration environmental factors that are likely to influence the sales volume. The degree to which the economic environment can affect sales varies depending on the nature of the product and the type of customer segment.
- Inflation is another important factor that would have an impact on the pricing strategies of the company. It also impacts the consumer's ability to pay for the company's products. If the inflation index shows a rise, a consumer is forced to cut down on his expenses (assuming his income does not rise proportionately). This may affect the demand for the company's product, particularly if the spending for the product is discretionary. The company may be forced to give discounts.
- The company also needs to examine how its sales have been varying for the past years. If the variations in the company's sales are in sync with the secular trends (economic performance and the overall sales for the product in the industry), it may be all right. However, if the company's sales show a substantial deviation, particularly below the market trend, the cause needs to be examined. For instance, this could indicate that the company is serving

NOTES

NOTES

a smaller market segment, or that there is something wrong with its marketing strategies.

- A repetitive pattern in the sales trend may also be valuable for the company. Traditionally slack and peak demand periods can be identified, and the company can prepare itself to meet sales challenges during these periods. Promotions can also be planned more accurately. Product life cycles, and seasonal fluctuations can also be gauged from the sales trends.

Sales forecasting should be done atleast on an annual basis. However, more frequent forecasts, particularly on a monthly basis, can be very helpful in understanding variations in the sales data.

Every forecast has some uncertainty built into it. The near term forecast is more likely to be accurate than a long-term forecast. Usually short-term forecasts are done for less than three months, intermediate forecasts are done for upto two years, and long-term forecasts are done for more than two years. Short-term forecasts are for planning, scheduling, procurement, staffing, logistics and inventory decisions. Intermediate forecasts are done for planning various marketing activities such as managing the salesforce, investments in plant and machinery, distribution decisions, and new product launch. Long-term forecasts are for strategic decisions, such as market expansion.

Forecasting Techniques

Several forecasting methods are available. A company should choose the one that is most suitable for it. It should also use accurate data to draw up the sales forecast. Additionally, the market conditions, resource availability and environmental conditions must be factored while making the forecast.

Either forecasts can calculate the total sales for a company or they can be more specific. Forecasts can be used to estimate sales per salesperson, per territory, per product, per branch or for each customer segment.

Approaches to Forecasting

Causal approach involves identifying various factors that cause sales variations. A few causative factors are uncontrollable, as they are exogenous in nature. Changes in socio-cultural factors, economic environment, economic conditions in the market and the competition are some such uncontrollable factors. However, all the variables in the company's marketing strategy are controllable factors. They can be monitored and changed by the company.

The non-causal method does not identify the causes of the fluctuations of sales. Therefore, the presumption is that the factors that caused sales fluctuations will continue to exist the next year as well. The method is not very scientific as it fails to analyze the reasons for the sales variations. In addition, it is very presumptuous on the part of businesses to assume that the same reasons will cause similar variations in the forthcoming period. Given the dynamic nature of business environments, it is unlikely to be so.

There are direct and indirect methods of estimating the total sales for a company. The indirect method first predicts the total sales of the industry, and the company's sales are a part of this whole. In the direct method, the company's sales are estimated directly. Of these estimations can be done by either causal or the non-causal methods.

There are fundamentally two types of sales forecasting methodologies:

In bottom-up forecasting, the market is divided into segments. Sales for each segment are estimated. These are then combined to arrive at the total forecasted sales. In top down forecasting, the sales potential is estimated first. Thereafter, sales quotas are divided for each marketing unit, and then the sales forecasts are done. The accuracy of bottom up forecasting depends on the truthfulness of the answers to buy intentions given by customers, while the accuracy of top down forecasting depends on the ability of the company to understand market conditions and environmental conditions. These two methods can be used in various sales forecasting techniques that can be broadly classified under qualitative and quantitative techniques.

Qualitative Methods

Qualitative methods are based on judgments of salespersons, experts and consultants. These methods do not use statistical tools to arrive at the sales forecast. The qualitative methods suffers from the obvious disadvantage of being driven by the opinions of individuals. However, it is particularly suitable when a new product is launched in the market and no past records are available. When customer tastes and preferences change dramatically, these methods may be very useful. Essentially, when there are far too many unknown variables, or there is lack of any specific information, qualitative methods are a good starting point for estimating sales.

Probability Assessment Method (PAM)

This method calculates total sales by assigning probabilities to various levels of sales. The sales are multiplied by the probabilities to arrive at total sales. The cumulative sales are estimated. The cumulative sales curve aids in forecasting.

The Program Evaluation and Review Technique (PERT) uses three estimates: optimistic, pessimistic and most likely scenarios in the future. An expected value is arrived at using these three estimates. The standard deviation of the expected value is calculated and the confidence interval is calculated.

The Delphi Technique essentially works on the basis of the opinions of experts. The experts arrive at the consensus median sales figure after a lot of deliberation. The experts do not arrive at the sales figure as a group. Rather they are isolated in order to avoid the pitfalls of group discussions, such as the presence of a dominating individual.

Visionary forecast is a method in which a renowned expert who possesses insights about the business arrives at a sales figure based on his expertise. Even though the method is quite unscientific, it is useful for small businesses.

NOTES

NOTES

Historical analogy is the method that analyzes the introduction and growth of a product that is similar to the one whose sales are being estimated by the company. The method requires the sales history for the past several years. One or more products that bear a close resemblance to the product being forecasted can be used in this method. The accuracy of the method depends on the accuracy of the past sales data of the other products.

Intention-to-buy survey is a method that estimates the intention of the customers in the company's target market to buy the product within a well-defined future time period. The product on offer is described to the customer in detail, and he is asked relevant questions. This is a survey-based method, which depends on the respondent's ability to comprehend the product, and the questions asked. Further, it relies on the honesty of the respondent while answering the questions. The greater the time lapsed between the survey and when the respondent bought the product, the less accurate the forecast will be. This is because the number of extraneous factors that can affect the intention of the customer to buy will increase.

Quantitative Methods

Time-series analysis and projection use past variations in sales to estimate future sale. The forecast identifies trends that form patterns to cause sales fluctuations. These variations could have been caused by seasonal fluctuations or cyclical patterns of the business. These predictable fluctuations, can be incorporated to yield a reliable forecast. Some random fluctuations could also exist in the past sales data. These make sales forecasting tougher.

Time series is generally used for intermediate and long-term forecasts. A sales history of at least five years is usually required to make good predictions. Though the time series does not really recognize the causative factors that lead to sales fluctuations, predictions are made based on past trends. Considerable data is required for making an accurate forecast. Preferably, at least two sets of data are required to ensure accuracy. Information is also required about a host of market variables that are likely to impact sales.

In simple-trend analysis, an extrapolation of the existing sales data is made. However, the simple extrapolation suffers from the disadvantage of oversimplicity. It just assumes that conditions would remain the same, and does not take into account sales fluctuations due to random occurrences that could have or will cause variations.

Moving average is a more sophisticated method of estimating sales. The future is assumed to be an average of the past, rather than a simple linear trend. It minimizes the impact of random variations, as it is an average.

The industry-survey method uses a survey-based methodology to forecast sales. The sales for the entire industry that manufactures the product is forecasted. A top-down forecasting method is used for this purpose.

A regression analysis can also be used to estimate sales. In a linear regression equation, a relationship is developed between sales as a dependent variable and an independent variable that causes fluctuations in sales. Since linear regression models are simplistic in nature, multiple regression modes are more accurate in predicting sales. A relationship is developed between sales as a dependent variable, and several independent variables, that are causative factors for variations in sales. This can be done by establishing a multiple regression equation based on past data. If the relationship between the independent variables and sales (the dependent variable) is found to be significant, the independent variable is a significant cause for fluctuations in sales. Therefore, the regression equation can be used to ascertain sales figures for the forthcoming time period.

Exponential smoothing is similar to the moving average method used in time series. It is however, a more refined version, since recent data is given higher weightage. More advanced models can incorporate fluctuations arising from seasonal and other factors.

Causal methods are particularly useful for developing sales forecasts when new products are being launched. In order to establish cause- effect relationships between variables, market surveys are necessary.

13.8.2 Dating Mining

The process of collecting and sourcing data related to the association, pattern, anomalies, changes from a large volume of data in the databases is called data mining. Data mining is one of the most popular tool used by different industries like marketing, retail, financial, and communication for understanding the customer interests and preferences. Data mining helps the companies to form a relationship between different factors related to the customers. This includes price, positioning and staff related factors to the customer profiles, competition and economic conditions. Such analysis reveals important metrics related to customer satisfaction, profits, sales. It also highlights a great deal of information about the transactional data of the customers.

Check Your Progress

14. What does media research do for the advertiser?
15. List one main component of a sales forecasting system.
16. What is the main advantage of sales forecast technique?
17. What does the causal approach to forecasting involve?
18. What is the visionary forecast method?
19. Which disadvantage does simple extrapolation suffer from?

NOTES

13.9 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

NOTES

1. Focus groups can be helpful in specific ways to:
 - (a) generate hypotheses about consumers and market situations
 - (b) suggest fresh and revitalised ideas.
2. Demographics research is concerned with gathering vital statistics about consumers – such characteristics as their age, income, sex, occupation, location, race, marital status, and education.
3. Some tasks of product research are:
 - (i) Development of a new product altogether.
 - (ii) Introduction of new functional features in the product.
 - (iii) Test marketing a new product.
4. A product is analyzed for what it has (merits), what it is (features), and what it does (functional utility and benefits drawn).
5. The company desire to maximise unit sales believing that a higher sales volume leads to lower unit costs and higher long-run profit. Price will be fixed at the lowest assuming the market price is sensitive. This is called market-penetration pricing.
6. A skimming price strategy may be used when a new product is substantially different from competitive offerings and the market is not price sensitive and when the marketer wants quick returns.
7. While designing the price structure of different product items, the firm should take into consideration the factors like (i) cost difference between (among) the items, (ii) customer evaluation of the different features and (iii) the competitor's prices.
8. Generally, direct questioning does not give correct answers for one of the two reasons: either the consumer does not know his true reasons or he knows but is unwilling to tell them to the researchers or gives illogical reasons.
9. From the standpoint of a market researcher, brand product identification is a total disaster. In response to direct questioning concerning purchase behaviour, all that the researcher gets is a major brand name-even when the consumer is buying "off-brands".
10. In depth distribution, research about a brand/product provides information related to its retail presence, market size and share, sales achieved, how well the competitor is selling and seasonality of demand.

NOTES

11. Media research attempts to eliminate waste in advertising by objectively analyzing the media available for promoting products and services.
12. The goal of advertising often centers on attempting to alter the attitudinal set for the brand in question within a given segment in the direction of those product characteristics which were considered to be salient.
13. The following are the main steps in formulation of advertising objectives and budget:
 - Setting up operational advertisement objectives.
 - Determining the tasks that have to be performed to attain these objectives.
 - Arriving at the advertising budget by totalling the costs of performing the tasks.
14. Media research attempts to evaluate and determine the most suitable and convenient media for the advertiser. It attempts to determine and assess the reach and impact of various media.
15. One main component of sales forecasting system is: Market Research Operations to get the relevant and reliable information about the trend in market.
16. The main advantage of the sales forecast technique is that it helps the company to plan its material and resource requirement for the coming year based on the anticipated sales volume.
17. Causal approach involves identifying various factors that cause sales variations.
18. Visionary forecast is a method in which a renowned expert who possesses insights about the business arrives at a sales figures based on his expertise. Even though the method is quite unscientific, it is useful for small businesses.
19. The simple extrapolation suffers from the disadvantage of oversimplicity. It just assumes that conditions would remain the same, and does not take into account sales fluctuations due to random occurrences that could have or will cause caused variations.

13.10 SUMMARY

- The two major strategies of consumer behaviour are exploratory and conclusive studies.
- Focus group interview is another popular technique for exploratory research. Focus groups generally bring together eight to ten people with similar

NOTES

backgrounds in a casual setting to apply the principles of group dynamics and free association to a marketing problem.

- In consumer research studies primary data can be collected either at one time or over a period of time. These research designs are referred to as a cross sectional and longitudinal, respectively and the two approaches have different purposes.
- B2B marketing research is the process of uncovering insights into the marketplace by surveying a representative sample of its participants.
- Product research is essentially concerned with satisfying the need of the consumer in the best possible way by giving him the most optimum product. Product research would set out to position the product on its attributes in a manner that the desired benefits are perceived by the customer.
- Price is the main element in the marketing mix that produces revenue through the sales. Achieving the long-run pricing objectives always depends on sound price policy as well as price strategy.
- A very high price is fixed for a new product at the initial stage. The main purpose is to make maximum profits from the initial demand.
- A skimming price strategy may be used when a new product is substantially different from competitive offerings and the market is not price sensitive and when the marketer wants quick returns.
- Price is to be modified according to the product which is the part of the product mix. Here the firm wants to increase the profitability on the total product mix. Pricing becomes difficult when the various products have demand and cost interrelationships.
- Businessmen expect (probable) return on their investment. Many well established companies follow the pricing objective in terms of return on investment (ROI).
- Motivation Research “is a term used to describe the application of psychiatric and psychological techniques to obtain a better understanding of why people respond as they do with products, advertisements and various other marketing situations.”
- Consumers are generally unaware of their own “true” reasons for specific brand and product selections.
- Marketing and distribution research is a key means to understand and analyze the marketing environment and thereby helping eliminate any weak links present.
- The advertising research is an application of marketing research aimed at the measurement of advertising effectiveness (how well an advertisement or

advertising campaign accomplishes its objectives) and ameliorate advertising efficiency (the most optimum use of the advertising budget and media-mix in implementing an advertising campaign).

- Advertising research with is concerned the measuring the attitudes of consumers to whom the advertisement is targeted. After measuring the attitudes, we attempt to determine the effect of advertising on these attitudes.
- Advertising message focuses on the ability of the advertisement to achieve impact and to project the desired message.
- The primary function of media research is to select the media most efficient for our ad campaign (media mix research). It is a part of media planning.
- Due to dynamic nature of market phenomenon sales forecasting has become a continuous process and requires regular monitoring of the situation.
- The record of previous year's sale is the most reliable basis as to future sales as the past performance is based on actual business conditions.
- The sales forecast preparation requires the availability of the sales turnover of the company for the past years, so that the company can understand the sales patterns.
- Forecasting helps the company to allocate sales quotas per region, territory, branch and salesperson. Further, it also allocates quotas for very customer segment, or individual customer (for large customer accounts).
- Either forecasts can calculate the total sales for a company or they can be more specific. Forecasts can be used to estimate sales per salesperson, per territory, per product, per branch or for each customer segment.
- Probability Assessment Method (PAM) calculates total sales by assigning probabilities to various levels of sales. The sales are multiplied by the probabilities to arrive at total sales.
- The industry-survey method uses a survey-based methodology to forecast sales. The sales for the entire industry that manufactures the product is forecasted. A top-down forecasting method is used for this purpose.

NOTES

13.11 KEY WORDS

- **Cross-sectional Study:** It is defined as a type of observational research that analyzes data of variables collected at one given point in time across a sample population or a pre-defined subset.

NOTES

- **Distribution Research:** It refers to the collection and analysis of information related to the sales of a product or brand and its distribution through various retail channels so as to enable the management make better decisions.
- **Advertising Research:** It is a systematic process of marketing research conducted to improve the efficiency of advertising. Advertising research is a detailed study conducted to know how customers respond to a particular ad or advertising campaign.
- **Vehicle Exposure:** It refers to the number of people and the type of people reading an average issue of the media. It also includes the frequency - the number of times each person reads an average issue.

13.12 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the major roles of conclusive research?
2. What is B2B marketing research?
3. List some factors that influence pricing.
4. Write a short note on skimming the cream pricing strategy.
5. Define advertising research.
6. What does the scope of advertising research include?
7. What does the study in advertising research measure?
8. Write a short note on product appeal research.
9. Write a short note on the importance of sale forecasts.
10. List two factors to be considered in a sales forecast.

Long-Answer Questions

1. Analyze the scope and sources of product research.
2. Discuss the issues before a company to pursue its objects through pricing policy.
3. Explain any three pricing strategies.
4. Elaborate upon the product mix pricing strategies.
5. Discuss the objectives of pricing.
6. Explain the reasons as to why direct questioning does not give correct answers.

7. Elaborate upon the classification of advertising research.
8. Analyse the factors to be considered in a sales forecast.

*Application of Marketing
Research*

13.13 FURTHER READINGS

Saravanavel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.

Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.

Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.

Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.

Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.

NOTES

NOTES

UNIT 14 RECENT TRENDS IN MARKETING RESEARCH

Structure

- 14.0 Introduction
- 14.1 Objectives
- 14.2 Marketing Information / System and Research
 - 14.2.1 Online Marketing Research
 - 14.2.2 Recent Trends in Marketing Research
 - 14.2.3 Research in Lifestyle Retail
 - 14.2.4 Marketing Research and Social Marketing
- 14.3 Rural Marketing Research
- 14.4 Trends in Services Marketing Research
- 14.5 Brand Equity Research
- 14.6 International Marketing Research
 - 14.6.1 Branding Research
- 14.7 Answers to Check Your Progress Questions
- 14.8 Summary
- 14.9 Key Words
- 14.10 Self Assessment Questions and Exercises
- 14.11 Further Readings

14.0 INTRODUCTION

Marketing research is an important activity undertaken by marketers to gather information regarding the target markets and customers in order to evaluate who they are, what their preferences are and so on. According to Philip Kotler, “A marketing information system is a continuing and interacting system of people, equipments, and procedures to gather, sort, analyze, evaluate, and distribute the pertinent, timely, and accurate information for use by marketing decision-makers to improve their marketing planning, implementation, and control.” Some recent trends that have appeared in marketing research over the years are automation, artificial intelligence, blockchain, mobiles and social media. The basic objectives of conducting market research include the assessment of the feasibility of a new business, developing strategies to adapt to the changing environment and evaluating customer needs to offer better services and products. This unit provides an analysis of the various types of research involved in marketing in addition to the recent trends in marketing research.

14.1 OBJECTIVES

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

- Discuss the meaning of Market Information System and Research

- Analyse the concepts of Rural Marketing Research and Brand Equity Research
- Describe the trends in Services Marketing Research
- Understand the dimensions of International Marketing Research

NOTES

14.2 MARKETING INFORMATION / SYSTEM AND RESEARCH

Each organization must organize the flow of marketing information to its marketing managers. The organizations are studying their managers' information needs and designing marketing information systems (MkIS) to meet these needs. Marketing Information System consists of people, equipment and procedures to later, sort, analyze, evaluate and distribute needed, timely, and accurate information to marketing decision makers.

Kinds of Marketing Information Systems

Basically, there are three kinds of systems which are adopted for this purpose:

1. **Marketing Control System:** It provides with the management information relating to the trends and problems in the market as also the marketing opportunities. The data provided cover numerous aspects of sales and provide the management with a basis to assess its performance in the past. Then this becomes a pointer for the future. The system is useful for controlling the marketing cost and finding out reasons for poor sales.
2. **Marketing Planning System:** It furnishes information required for future planning of products in a most convenient and intelligible form. It provides necessary materials even to decide alternative products in addition to the existing ones. The problems handled by this system are: Sales forecasting, Promotional planning, Credit management, etc.
3. **Marketing Research System:** This is useful for analyzing and solving current marketing problems. It measures characteristics of different types of customers and also their behaviour under different conditions. Marketing research handles problems concerning advertising, price, etc.

In short, marketing information system is a set of procedures and methods for the regular, planned collection, analysis and presentation of information for use in making marketing decisions.

Marketing research is an integral part of the marketing information system. It is both a science and an art. It is a science because there is systematic application of the principles of research. It is an art because it tells us how to solve day-to-day and specific problems of management, particularly problems of marketing management. The marketing information system concept is illustrated in Figure 14.1

NOTES



Fig. 14.1 Marketing Information System

The marketing managers, in order to carry out their analysis, planning, implementation and control responsibilities, need information about developments in the marketing environment. The role of the MkIS is to assess the manager's information need to develop the needed information and distribute the information in a timely fashion to the marketing managers. The needed information is developed through internal company records, marketing intelligence activities, marketing research and marketing decision support analysis.

The most basic information system used by marketing managers is the internal records system. Included are reports on orders, sales, prices, inventory levels, receivables, payables and so on. While the internal records system supplies the 'results' data, the marketing intelligence system supplies 'happenings' data. A marketing intelligence system is a set of procedures and sources used by managers to obtain their everyday information about pertinent developments in the marketing environment. Then, marketing managers commission formal research studies of specific problems and opportunities. Lastly, the marketing decision support systems, which consists of statistical techniques and decision models, assists marketing managers in making better decisions.

Marketing Research and Marketing Information Systems

Marketing Information System is a structured, interacting complex of persons, machines and procedures designed to generate an orderly flow of decision-making in specified responsibility areas of marketing management. This definition indicates the interdependent activities associated with the collection of marketing information, both from internal and external sources. It also shows that such information is collected to facilitate decision-making in different areas of marketing management.

Why we need a Marketing Information System (MkIS):

- **Fast Decisions:** Today, executives have less time to take the decisions, and are expected to react fast to changes, and manage new product launches.

- *Complexity of Marketing*: Marketing as a function has become more comprehensive and broader. Transnational and global marketing makes it imperative to have good information system.
- *Era of Shortages*: Shortages of resources like energy and other raw materials make it compulsory for management to keep only viable products in its product line.
- *Consumer Awareness*: Consumer expectations are high and to live up to these, managements need information.
- *Information Explosion*: Makes it necessary to process it so that best use is made of it. Thus, a good marketing information system should determine the information needs of the organization and generate and process such information on a continuing basis. It should also provide for its storage so that it can be used when required.

NOTES

14.2.1 Online Marketing Research

As the name suggests, this refers to the method of carrying out the marketing research process including the data collection over the internet. The bases of the research process like the collection of information related to the preferences, taste and nature of the target market remains the same, except the sources of information and their collection is not done in the field but from internet sources. The sources and nature of data even here can be qualitative or quantitative. While online communities and forums, or tools like video ethnography, or social media may be examples of sources of qualitative data; mobile survey and app surveys are called quantitative data sources.

Online marketing research, as you have seen before is extremely important because it is used for preparing plans and goals, and that affects the strategies for sales and revenue and also the product planning processes and many more areas. Online marketing researches are gaining currency because internet in the recent decades has become an indispensable part of our lives. A majority of the time of a majority of the population is spent online. These habits, preferences, interactions, opinions, etc., are crucial for marketing processes. Online marketing is done to gain knowledge about the target customer, learn about the behaviour of the customers and learn about the potential areas for making revenue.

14.2.2 Recent Trends in Marketing Research

The following are examples of some of the trends that can be seen in the field of marketing research:

- **Automation**: In many different fields, the dominance of automation has been increasing. Even, in the marketing areas, machines are taking over several activities. Some of the areas of marketing research where automation is being observed include data collection and processing. Automation is being used as it is accurate, fast and is suitable for such tedious tasks. It also

NOTES

helps in easing the process of analyzing the findings through crisp and objective numbers.

- **Mobile:** Mobiles, especially in the last ten-fifteen years have become one of the most important tools of communication. Newer technologies are making information available at the fingertips, marketing can benefit from this and the usage data, surveys, viewing counts, app downloads, watch hours, etc., and other data can be used in marketing research. Additionally, advanced technologies like virtual and augmented reality will bring even more surprising ways of using mobiles.
- **Artificial Intelligence:** Artificial intelligence and its development has seen many upgrades over the recent years. Many researchers are focusing on making artificial intelligence more accurate. It is being used for sale and customer service already in many companies like Amazon. Further, surveys and questionnaire development process and analysis would gain majorly from artificial intelligence technologies.
- **Blockchain:** The field of marketing research is increasingly incorporating blockchain technologies to take the benefit of lowered frauds. Blockchain is nothing but an online ledger which duplicates and distributes data across the entire network of computer systems on the blockchain. It uses safety measures like new ID verification.
- **Social Media:** Marketing research is now increasingly making appearance in people's social media feed. This is probably the area which reflects individuals and their preferences in the most authentic and direct manner.

14.2.3 Research in Lifestyle Retail

Lifestyle retailing is described as the policy of tailoring a retail offer, or a portfolio of retail offers, closely to the lifestyles of specific market segment. Using lifestyle data is very helpful in determining a retailer's core customers and the composition of their trade area. Lifestyle research data provides another way of looking at the characteristics of a population. Lifestyle segmentation systems combine attributes of density (urban to rural location) with attributes of demographics to define a given lifestyle.

14.2.4 Marketing Research and Social Marketing

Social marketing is the use of commercial marketing principles and techniques to improve the welfare of people and the physical, social and economic environment in which they live. It is a carefully planned, long-term approach to changing human behaviour. A social marketing program has as its core the wants and needs of its consumers. These are determined through market research methods that aim to learn as much about the target audience and how it thinks, feels and behaves in relation to the issue the program is addressing. These methods include quantitative research, such as a knowledge, attitude and behaviour (KAB) survey, which reveals

how many people think or do something. Qualitative research, on the other hand, provides insight into why people think or do what they do, through techniques such as focus groups and individual interviews.

Check Your Progress

1. How is a marketing research system useful?
2. What is a marketing intelligence system?
3. Why is online marketing research important?
4. Define lifestyle retailing.

NOTES

14.3 RURAL MARKETING RESEARCH

Rural marketing is identifying and serving the needs of consumers living in villages. Some of these villages are located near cities and they enjoy facilities like electricity, running water and communication infrastructure. Some other villages are located in remote areas and modern facilities have not yet reached them. The primary source of income for rural consumers is agriculture.

Marketers have been content with treating rural consumers as less sophisticated than their urban counterparts and have been trying to serve them with less sophisticated products from their portfolio. It is a wrong strategy. Rural consumers live differently, their environmental conditions are different, and their source and pattern of income is different. Rural markets are different from urban markets in significant ways. It is important that marketers understand the living conditions of rural consumers before attempting to sell to them. Sadly, this cannot be done from the company headquarters. Marketers have to stay in rural areas and live rural life to understand nuances of rural consumer behaviour.

Researching Rural Markets

The researcher will have to live in the rural market for a reasonably long period. He will have to understand rural life in its real settings. He should have the sensibilities to notice the way they use products and services and how they avoid using them, and how they substitute one product for the other. He has to participate in their marriages and squat on the ground to have community meals with them on ceremonial occasions.

The big problem with rural marketing is that people who understand the nuances of rural consumer behaviour are not articulate or powerful enough to influence decision makers sitting in distant headquarters. And people who are powerful and articulate rarely have the sensibilities to understand rural consumption and consumers. Till the rural sensibilities and corporate power are combined in an individual, rural consumers will suffer innate policies and programmes of urban marketers.

14.4 TRENDS IN SERVICES MARKETING RESEARCH

NOTES

The growing importance and criticality of service sales and marketing to the economy through its contribution to the GDP is well documented. Additionally, the stage of development of a country could be directly correlated to the extent of evolution of service sales and marketing in that country. Thus you can see the service industry contributing to about 75 per cent to the GDP of developed countries whilst the corresponding contribution from lesser developed countries is significantly lower. This evolution in services marketing can be clearly understood by analysing one of the largest businesses in the world—the automobile business.

In the automobile business, service sales and marketing evolved as a means to support product sales initially. The focus of Original Equipment Manufacturers (OEMs) was to build vehicles which due to their poor quality broke down forcing customers to return to the automobile dealership for repairs ‘under warranty’. ‘Order taking’ was the principal activity of these early days of ‘service advisors’, and the main issue was in arranging for suitable capacity for enabling timely repairs. In this era, dealers ruled the roost by refusing to service vehicles not sold by them and insisting on appointments to be booked in advance for servicing/repairing vehicles sold by them.

As competition increased, the service aspect gained importance through the incorporation of the concept of ‘scheduled maintenance’. This concept stressed the proactive approach for providing service before a breakdown happened, in a non-stressful atmosphere and with better planning, leading to a win-win solution for all concerned. However, this practice led to a situation where a mindset towards earnings from maintenance crept in and the ‘service’ aspect was given less importance. In fact, in extreme cases, this led to dealerships instituting a system whereby reminders were sent to clients for routine maintenance primarily with an objective of increasing revenue from servicing. However, with further intensifying of competition (especially from the Japanese producers), OEMs started paying attention to get the quality ‘first time right’ and therefore the importance and scope for repair/maintenance along with the attendant revenues started declining. This situation catalyzed two distinctive actions:

- Success with the ‘first time right’ manufacturing process enthused the OEMs to start offering extended warranties and soon warranties for 100,000 km or five years (which on an average translated to half the lifetime of a vehicle) started getting offered. This therefore spelled the era where competitive advantage was attempted to be gained through service offers as the product differentiation lost its edge.
- Secondly dealerships which had started getting used to significant income through repair and warranty, servicing started offering services which

were focussed in enabling to provide ‘enhanced driving experience’ to its customers. This was done with the intent to offset the reducing revenue for repair and warranty and in a way is leading towards an era where the true spirit of servicing is coming of age.

The aforementioned example clearly exemplifies the growing inevitable evolution and importance of services marketing through distinctive steps as:

- Services marketing commences as a product sale support activity
- Services marketing gradually gains importance, as product differentiation reduces
- Services marketing evolves into a core sales and marketing activity on its own right

The day is not very far off when the entire ‘product dimension’ of this business may be substituted by the ‘transportation service’ approach and manufacturers start marketing the ‘transportation service’ to individual—what is being witnessed today in the ‘fleet sales’ segment. This evolution of services marketing can similarly be traced by a study of the evolution of literature on the subject. In the marketing field, service marketing has emerged as a discipline of well-established research in the past two decades or so. In fact, service appears to be gaining ground over goods, as the fundamental for economic exchange. The evolution of services marketing has been traced by Fisk et al. (1993) into three broad stages:

- **Crawling out (1953 to 1979):** This initial stage witnessed the attempts made by authors to distinguish services from goods. This early phase was drawn out as the authors had an uphill struggle in introducing service as a fundamental provision for economic exchange. The distinctive IHIP (Intangibility, Heterogeneity, Inseparability and Perishability) characteristics of services were identified at this stage and this effective differentiation led to the faster growth of this field subsequently.
- **Scurrying about (1980 to 1985):** Once the basic battle for differentiating services from goods was won in the first stage, this stage witnessed the implications of these differences between services and goods. Other aspects covered in this stage were:
 - o Developing insights for marketing practices in the service sector
 - o Classification of services
 - o Marketing treatment for each classification
 - o Impact of the IHIP characteristics on strategies, marketing practices and problems
 - o First forays into ‘service mapping’, ‘service design’ and ‘service encounter’
- **Walking Erect (1986 to 1993):** In this stage, services marketing established itself as a substantive field within the marketing discipline.

NOTES

NOTES

This maturity literature started covering the wide heterogeneity of the subject by exploring:

- o Service design
- o Service experience
- o Control of intangible processes
- o Supply and demand issues especially where service capacity constraints prevailed
- o Organizational structure issues arising from handling the above areas
- o Service quality and its measurement alternatives

Thus, the aforementioned evolution in service literature clearly mirrors the growing sophistication of this field within the marketing discipline and clearly highlights the increasing importance of services in propelling economic activity in these modern times. In the Indian context, the actual facts on the ground also reaffirm this reality as exemplified by the enormous impact of ‘service economy’ (based on IT and ITES services) on the GDP.

14.5 BRAND EQUITY RESEARCH

A brand is an intangible asset for an organization. The concept of brand equity originated in order to measure the financial worth of this significant, yet intangible entity.

Brand equity is the value and power of the brand that determines its worth. The brand equity can be determined by measuring:

- The price premium that the brand charges over unbranded products;
- The additional volume of sales generated by the brand as compared to other brands in the same category and/or segment;
- The share prices that the company commands in the market (particularly if the brand name is the same as the corporate name, or customers can easily associate the performance of all the individual brands of the company with the financial performance of the corporate);
- Returns to shareholders;
- The image of the brand on various parameters that are deemed important;
- The future earnings potential of the brand;
- Or a combination of the above methods. Some methods of measuring brand equity involve the formulation of a multiplier by using a combination of the above methods. Such multipliers as brand strength or brand esteem can be determined by combining several variables to ultimately arrive at the brand equity.

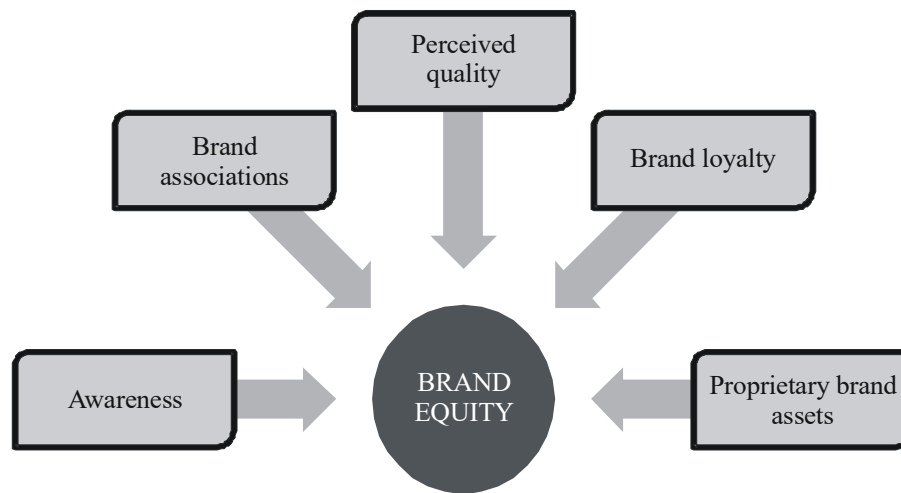


Fig. 14.2 Elements of Brand Equity

Brand equity comprises the following elements:

Awareness

Awareness of the brand name among target customers is the first step in the equity building process. Awareness essentially means that customers know about the existence of the brand and can also recall what category the brand is in. The lowest level of awareness is when the customer has to be reminded about the existence of the brand name, and that it is being a part of a particular category. In aided recall, the customer can recognize the company's brand from among a list of brands in the category. In unaided recall, the customer himself mentions the company's brand. The highest level of awareness is when the first brand that the customer can recall upon the mention of the product category is the company's brand. This is called top-of-mind recall.

Awareness of the name acts as an anchor to which everything else about the brand is linked, much like the name of a person acting as an anchor for tying all associations about him. Building awareness involves making the brand visible to the relevant target audience by various promotional methods such as publicity, sponsorships, events, advertising, instigating word-of-mouth promotion, etc.

Brand Associations

Anything that is connected to the customer's memory about the brand is an association. Customers form associations on the basis of quality perceptions, their interactions with employees and the organization, advertisements of the brand, price points at which the brand is sold, product categories that the brand is in, product displays in retail stores, publicity in various media, offerings of competitors, celebrity associations and from what others tell them about the brand. And this is not an exhaustive list.

NOTES

NOTES

Consumers add to brand associations with each and every interaction they have with the brand. All these associations are not formed only due to their interactions with the organization. Many associations are formed from what others tell customers about the brand. It is absolutely crucial that the company plan each interaction with every customer and relevant others (media, shareholders, employees, government) so as to eliminate even the slightest chances of any negative associations that can emanate from any of these sources.

Associations contribute to brand equity, as strong, positive associations induce brand purchases, besides generating good word-of-mouth publicity. Such associations can also help the company in leveraging the brand, create strong barriers to entry for competitors, give trade leverage to the company and enable the company to achieve differential advantage.

Perceived Quality

Perceived quality is also a brand association, though because of its significance, it is accorded a distinct status while studying brand equity. Perceived quality is the perception of the customer about the overall quality of a brand. In assessing quality, the customer takes into consideration the performance of the brand on parameters that are important to him, and makes a relative judgment about quality by assessing competitor's offerings as well. Therefore, quality is a perceptual entity, and consumer judgments about quality vary.

Quality perceptions influence pricing decisions of companies. Better quality products can be charged a price premium. Quality is one of the main reasons for consumer preference for a brand in any product category. Thus, superior perceived quality can also be used to position the brand.

Brand Loyalty

A customer is brand loyal when he purchases one brand from among a set of alternatives consistently over a period of time. In the traditional sense, brand loyalty was always considered to be related to repetitive purchase behaviour. For some products such as purchasing a house or an automobile, repetitive purchase behaviour may not occur. In these situations, attitudinal brand loyalty, i.e., consumer feelings about the brand that was purchased, and their inclination to recommend the brand to others are measured. Brand loyalty is usually rated as the most important indicator of brand equity because loyalty develops post purchase and indicates a consistent patronage by a customer over a long period of time whereas all other elements of brand equity may or may not translate into purchases.

Brand loyal customers form the bedrock of a company. Higher loyalty levels lead to a decrease in marketing expenditure as such customers act as positive advocates for the brand. Besides, a company can introduce more products in its portfolio that are aimed at the same customers at less expenditure. It also acts as a potential barrier to entry for new players and gives time to the company to respond to competitive threats. The bargaining power of the company with the

trade channel members is stronger when there are many loyal customers who would only buy the company's brand. In this case, the retailer merely distributes the manufacturer's products.

Other Proprietary Brand Assets

Proprietary assets include patents, trademarks and channel relationships. These assets are valuable as they prevent competitors from attacking the company and prevent the erosion of competitive advantages and loyal customer base.

All activities of the firm determine brand equity. These activities may enhance or diminish the brand value. Activities that are synchronous with the overall vision for the brand enhance equity, and any activity that goes against this overall vision reduces brand equity.

Leveraging Brand Equity

Companies that build brand equity capitalize on such strong brands by using them to launch new products in other categories, or serving other customer segments in the same category, or serving the same customers in the same category better. The main purpose of using the same brand name is to take advantage of the value and power that the brand commands, rather than building a completely new brand, which would entail a huge expenditure, and would take time.

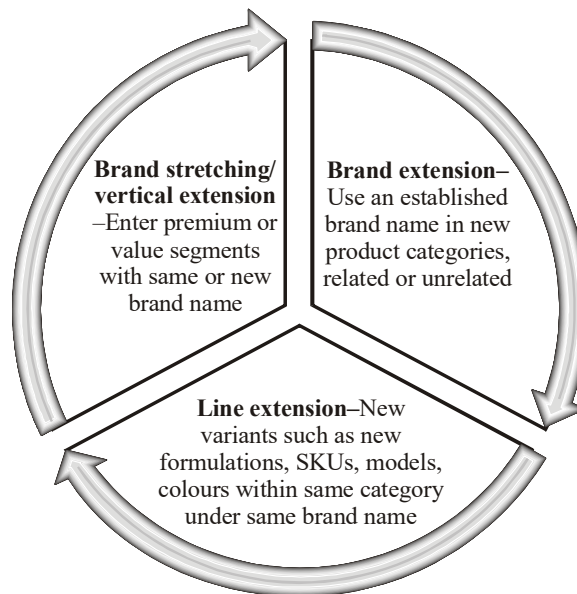


Fig. 14.3 Leveraging Brand Equity

Brand Extension

Brand extension is the use of an established brand name in new product categories. The category to which the brand is being extended can be related or unrelated to the existing product categories.

NOTES

NOTES

Brand extension in unrelated markets may result in loss of credibility if a brand name is extended too far. A company has to find out the product categories in which the established brand name will work and the product categories in which it will not work. To be able to do that, the company has to find as to why the brand name is successful in its current business. It may find out that customers in the current business desire luxury and exclusivity, and the brand is correctly positioned. It is offering luxury and exclusivity. If the customers of the new business also desire luxury and exclusivity, the brand name will work in the new business. If values and aspirations of the customers of the new business match those of the original business and if these values and aspirations are embodied in the brand, it is likely to be accepted by customers in the new business.

Advantages of brand extension in releasing new products are that it reduces risk and is less costly than alternative launch strategies⁹. Customers appear to attribute the quality associations they have of the original brand to the new product¹⁰. An established name enhances consumer interest and willingness to try the new product bearing the established brand name¹¹. Since the established brand is already well known, the task of building awareness for the new product is eased. Advertising, selling and promotional costs are reduced. There is likelihood of achieving advertising economies of scale since advertisements for the original brand and its extensions reinforce each other.

But brand extensions that offer no functional, psychological or price advantage over rival brands in the new category often fail. There is also the danger that management may not provide enough funds for the launch believing that the spin-off effects from the original brand name will compensate. This can lead to low awareness and trial. Also, bad publicity for one product affects reputation of other products under the same name. A related problem is the danger of the new product failing or generating connotations that damage the reputation of the core brand.

A major test of any brand extension opportunity is to ask if the new brand concept is compatible with the values inherent in the core brand. Brand extension is not viable when the new brand is developed for target customers who hold different values and aspirations from those in the original market segment.

Line Extension

New variants such as new product formulations, flavours, SKUs (sizes), models or colours within the same product category are launched bearing the established brand name.

Line extensions can be useful to reach out to new customer segments who seek new benefits, hitherto not being offered by the brand in the category. For instance, launching a new shampoo variant aimed at consumers seeking solution for a dry, itchy scalp under the existing shampoo brand name makes sense. Line extensions are also useful in reviving consumer interest in a dull product category. In some categories, variety is a desirable attribute for consumers.

However, line extension leads to managerial focus on minor modifications, packaging changes and advertising rather than on real innovations. Cannibalization can also occur, i.e., the new brand variants gains sales at the expense of the established variants of the same brand.

Brand Stretching or Vertical Extensions

The company may sense an opportunity in the premium or/and value segments of the market, besides its existing markets. The company enters into these premium or/and value segments as well. These movements are called upscaling and downscaling respectively.

Companies engage in brand stretching because:

- Existing markets may be saturated
- They sense new opportunities in the premium or popular segments of the market
- Competition in the existing segments is intense

Brand stretching strategy can be practised in three ways:

- Upscaling or downscaling the entire brand.
- Introducing a totally new brand in the upscaled or downscaled position.

Both the above techniques represent extremes. Almost all companies tread the middle path.

- Treat the original brand as the 'Mother Brand' and introduce its offshoots in the form of 'sub-brands'. These sub-brands can either be added as:
 - (a) Prefixes or suffixes.
 - (b) Describe the calibre of the stretched brand by using appropriate numbers/symbols/text abbreviation/ words/names.

When upscaling is done using existing brands, it might not be successful as customers in the premium segments may not accept the brand that was present in the popular segment earlier. Through upscaling, an existing brand may not necessarily fail, the customers in premium segments will not like to associate with popular brands. To overcome such associations is an onerous task. Using sub-brands with clearly distinguishable personality and values, or introducing a completely new brand may be necessary in this case. A new brand is the safest option for upscaling as it does not carry any baggage of associations with existing market segments. However, it is the riskiest and most expensive option as well.

Downscaling, on the other hand, can be accomplished easily with an existing brand, as customers in the popular segment are glad to buy a brand that is hitherto aspirational, due to its presence in the more

NOTES

NOTES

upscale market segment. But the current customers of the premium brand may not like to see their brand being associated with a downscale market and will in all probability, stop patronizing the brand. They will switch to some other premium brand. The company has to decide if the downscale segment is really big enough to take the risk of losing customers of the upscale segment.

Most companies have been enamoured with the prospect of serving mass markets and have downscaled premium brands that they had so painstakingly built for decades, sometimes even centuries. They have lost it all in a few years. And the mass market for which they took this hit, is so littered with entrenched competitors that after the initial hype of customers being able to own brands that they never thought they would own, the erstwhile premium brands have become just one of the competitors. And since the downscaled brand does not have experience of doing business in the mass market, it has even been muscled out of the mass market by the big fish here. Mass markets operate by a different set of machinations of scale, quality and price, which premium brands find very difficult to master. Parker pens and many other iconic brands have followed the above script and dug their own graves. Mass market has been the graveyard of many illustrious brands.

Using sub-brands are also likely to have similar effects. The mother brand, still in the more premium segment, can face repercussions if it is not appropriately distinguishable from the sub-brand in the popular segment. Many consumers from the premium segment can move to the popular segment, thus resulting in losses. A completely new brand for downscaling is a better option than using sub-brands for a firm that wants to be active in both the premium and the popular segments.

Check Your Progress

5. What is the big problem with rural marketing?
6. List any two aspects covered in the scurrying about (1980-1985) stage.
7. When is brand extension not viable?
8. Why is it easy to accomplish downscaling with an existing brand?

14.6 INTERNATIONAL MARKETING RESEARCH

Most international marketing decisions in international business are decided with the help of marketing research. Almost all tactical decisions require some amount of international market research. Researches have to be elaborate and effective

for making any decision. However, the Internet can be a good source for obtaining information that does not require validation of market. It is also a good source of doing a quick search. The Internet has emerged as a major source of providing the needed information for making decisions.

International marketing research facilitates marketers to make more informed decisions. International market research provides a clear picture of the market and based on the market situations, marketers can tell what is occurring or what is likely to occur and offer alternatives from which choices can be made.

An effective international market research may facilitate marketers in identifying multiple options for introducing new products /services or options of entering new markets. International marketing decisions may prove less risky if they are supported by market research.

But marketers, especially in the international scenario, should be aware that marketing research is only one of the several tools for making marketing decisions. Marketing research is concerned with the factors that are directly involved in the marketing of goods and services, and it includes the study of the effectiveness of the marketing-mix, advertising strategies, competition and consumer behaviour. There are many decisions that are based on marketing research (analyzed via quantified data collected on vital market indicators). These are:

1. Product-Market Decisions

- Identifying market characteristics
- Product-mix research
- Determining product sustainability
- Innovative product range
- Preferential and profitable positioning of products
- Distribution analysis
- Pricing strategies impact analysis
- Product testing—pilot studies
- Market share analysis
- Short and long range forecasting for price and demand of the product
- Sales trend analysis
- Competition pattern
- Consumer-behaviour analysis in reference to price, product-mix and comparative advantages over other products
- Assessment of impact advertising
- Analysis of gender preferences of products

NOTES

NOTES

2. Country Decisions

- Entry Decision
- Financial Risk
- Legal risk
- Opportunities
- Production and marketing costs

However, a good marketing research approach needs to be characterized by the following qualities:

- Scientific method
- Originality and creativity
- Potential to use multiple methods for cross checking the emerging results
- Interdependence on analytical models and data sets
- Cost of research

Approaches to International Market Research

Although conducting market research at the international level is important for gaining knowledge and for making marketing decisions, there is also the need to understand its limitations. All types of researches, whether for business, medical science, or the government are important for making decisions. There is also the risk that the results of these researches are wrong.

In one example, it has been noticed that between 1999 and 2002, many dot com companies failed because the information that was collected was not accurate and correct. This information was also not based on any scientific test. This happened because information was collected without the necessary controls. It has been observed in many cases that decisions may prove false and problematic if they are based on false/incorrect information. It is equally important to check if the sample sizes that many companies take in their surveys adequately cover most customers.

For obtaining the right and relevant results there is always the need to exercise strict control on how the information is collected. There are now many scientific methods that have been used in conducting market research. It is very important to know which method has been used to conduct research in international markets. But sometimes very structured and rigid controls may have a negative effect and, therefore, researchers should be very careful in adopting the right method to do the market research.

For example, in researches that require a sample of a large population of customers, it becomes difficult to control these kinds of sample sizes. Researchers should keep certain controls in place to define how many are in the sample as well as how the research is to be carried out (e.g., research design should be the same for all).

Large sample sizes result in an increase in cost and time. It is, therefore, essential to assess the benefits from the outcomes of the research in terms of the time and cost involved in carrying out the research. The trade-off for getting more relevant results is the increase in cost and time needed to carry out the research. So when it comes to doing research, marketers have to determine the balance between relevancy of the information obtained and the costs involved in carrying out the research.

A good market research is one whose results are strong indicators of what is happening now or might happen in the future. Results of a good market research should be relevant to the market situation.

The following step-by-step approach is useful for companies to do international market research. Major steps include screening potential markets, assessing the targeted markets, and drawing conclusions.

1. Screen Potential Markets

Step 1. Current export statistics are required to be obtained. These statistics are helpful in knowing the status of export markets. The Ministry of Commerce, Government of India, provides these statistics in a published format. Trade statistics are also available at the Reserve Bank of India, EXIM Bank and also various Chambers of Commerce in India.

Step 2. There is the need to identify five to ten large, emerging and fast-growing markets in which the product is likely to be marketed. Its performance over the past three to five years is required to be assessed.

Step 3. The fast-emerging markets may be identified. These provide ample opportunities to export. One may face lesser competition because in these countries markets have just opened up. Therefore, they offer good opportunities to enhance growth rates and acquire market share.

2. Marketing Information

Marketing decisions depend on marketing information. This is widely accepted as essential in domestic marketing situations, but its collection and interpretation is more problematic and, thus, less commonly used in international marketing. The following are the areas of substantial differences and, thus, potential problems between the domestic and international research scopes:

- New parameters—such as currencies, taxes, documentation and transportation systems.
- New environments—such as cultural, social, legal and political systems and ideologies. If a company is unaware of the differences that exist between home and foreign markets (for example, differences in culture, transportation systems or competition), it may not spend money gathering information on these factors.

NOTES

NOTES

- Number of factors involved—managers are faced with an increase in the complexity of decisions caused by the number of markets being entered, and the interaction of the elements under investigation.
- Broader definition of competition—competitive activity in developed domestic markets may be faced with established brands within a category. However, while going international, there may be other competitive substitutes; for example, soft drinks manufacturers might face competition from other beverages like tea, coffee, milk, drinks or fruit juices.

These factors interrelate to add to the complexity of understanding foreign markets.

International marketing managers need both feasibility data used to decide whether to internationalize, which markets to enter, how to enter and what to offer, and operational data, used for a range of strategic, tactical and operational marketing decisions once present in the market.

Feasibility and operational data can be gathered and analyzed within an international marketing information system, which harnesses data and information from both internal and external sources and on a continuous and ad hoc basis.

Marketing information systems should provide a system or process for gathering, analyzing, storing and retrieving relevant information about the company, its customers, its channels and its competitors, and the environment for actual and potential changes, and using the data to identify areas where action can improve competitive performance.

Data and information from marketing information system should be available to both field and head office staff, when and where required, and in a format which aids decision-making. Typically, this involves relational databases (those which can be integrated) between parts of the organization and sometimes also with its suppliers and customers. This flow of information will often use EDI (Electronic data interchange).

Reasons to Undertake Market Research

There are three major reasons to undertake market research.

(1) Describe what is happening (Descriptive Research)

Descriptive research is a widely used approach of doing research on the international market. This is also the most popular form of market research in international markets. Descriptive research focuses on providing an accurate description and/or explanation for something that is happening in the market. For example, “What segments/age groups are buying a particular brand?”, “What is a product’s market share in a specific geographic region?”, “How many competitors is the company facing?” are some of the issues dealt with in descriptive research.

(2) Exploratory Research: Examine something that is mostly unknown

To gain an insight into any issue, marketers do exploratory research. The basic difference between exploratory and descriptive research is in the research design. The exploratory approach in a research is an attempt to get information about some topic that is not well understood by the marketer. For example, a marketer may know about a new Microsoft web programme and technology for e-marketing a product, but he may not know much about the technology. Exploratory research works well when the marketer does not have an understanding about the topic or the topic is new, and it is hard to pinpoint the research direction.

(3) Causal Research: See how something affects something else

In this form of research, the marketer establishes a relationship between dependent and independent variables. This approach is used to test and understand different marketing scenarios. This is also a process of conducting an experiment. The design of causal research is highly structured and controlled. This kind of design is required for conducting experiments so that other factors do not affect those being studied. If causal research is performed well, the marketer may be able to use results to forecast what might happen if the changes are made. But in assessing markets, this kind of research should be used cautiously.

NOTES

14.6.1 Branding Research

A brand is a symbol of a sign which is adopted by a manufacturer for his product. It is a mark which is used by an organization while marketing a product. A brand is a device by which personality is given to a product. It creates an image. It is an attempt to isolate a part of the market by creating a brand image. People remember a product by its brand. Branding is a commonly accepted technique for product differentiation. Product research is concerned with branding because a brand is a deciding factor in the marketing of that product. Brands are the passports to exclusive or protected markets.

Further, research on branding is very much essential. Branding is a major issue in product strategy. Developing a branded product needs a great deal of long-term investment spending especially for advertising, promotion and packing. On this reason, the skill of marketer lies on his ability to create, maintain, protect and enhance consumers' product-loyalty. In other words, a brand is essentially seller's promise to consistently deliver a specific set of features, benefits and services to buyers. The best brand convey a warranty of quality but at the same time it is a more complex symbol which specifies attributes, benefits, values, culture, personality and user. Branding can be conducted after consulting a sample of customers of the company's existing products or potential customers of the new product. They may be asked whether they like this new brand name or not. The information is analysed and

NOTES

brand names finalised for the product on that basis. Various tests which can be conducted to test the branding policy soundness are: (1) *association tests* – to know that what images come to the respondent's mind after bearing about the present brand name; (2) *learning tests* – to collect information on ease in pronouncing the brand name; (3) *memory tests* – to know whether the name is easy to remember; and (4) *preference tests* – to know that which one of several names are preferred and on what basis. The decision is also to be taken on the basis of data whether the individual or family brand names would be successful and what strategies would be the best.

Check Your Progress

9. Which ministry provides current trade statistics?
10. What does descriptive research focus on?

14.7 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Marketing research system is useful for analyzing and solving current marketing problems. It measures characteristics of different types of customers and also their behaviour under different conditions. Marketing research handles problems concerning advertising, price, etc.
2. A marketing intelligence system is a set of procedures and sources used by managers to obtain their everyday information about pertinent developments in the marketing environment.
3. Online marketing research is extremely important because it is used for preparing plans and goals, and that affects the strategies for sales and revenue and also the product planning processes and many more areas.
4. Lifestyle retailing is described as the policy of tailoring a retail offer, or a portfolio of retail offers, closely to the lifestyles of specific market segment.
5. The big problem with rural marketing is that people who understand the nuances of rural consumer behaviour are not articulate or powerful enough to influence decision makers sitting in distant headquarters. And people who are powerful and articulate rarely have the sensibilities to understand rural consumption and consumers.
6. Some aspects covered in the scurrying about stage were:
 - (a) Developing insights for marketing practices in the service sector
 - (b) Classification of services
7. Brand extension is not viable when the new brand is developed for target customers who hold different values and aspirations from those in the original market segment.

8. Downscaling can be accomplished easily with an existing brand, as customers in the popular segment are glad to buy a brand that is hitherto aspirational, due to its presence in the more upscale market segment.
9. The Ministry of Commerce, Government of India, provides current export statistics in a published format.
10. Descriptive research focuses on providing an accurate description and/or explanation for something that is happening in the market.

NOTES

14.8 SUMMARY

- Marketing Information System consist of people, equipment and procedures to later, sort, analyze, evaluate and distribute needed, timely, and accurate information to marketing decision makers.
- Marketing information system is a set of procedures and methods for the regular, planned collection, analysis and presentation of information for use in making marketing decisions.
- The role of the MkIS is to assess the manager's information need to develop the needed information and distribute the information in a timely fashion to the marketing managers.
- Marketing Information System is a structured, interacting complex of persons, machines and procedures designed to generate an orderly flow of decision-making in specified responsibility areas of marketing management.
- The bases of the research process like the collection of information related to the preferences, taste and nature of the target market remains the same, except the sources of information and their collection is not done in the field but from internet sources.
- Marketing research is now increasingly making appearance in people's social media feed. This is probably the area which reflects individuals and their preferences in the most authentic and direct manner.
- A social marketing program has as its core the wants and needs of its consumers. These are determined through market research methods that aim to learn as much about the target audience and how it thinks, feels and behaves in relation to the issue the program is addressing.
- The big problem with rural marketing is that people who understand the nuances of rural consumer behaviour are not articulate or powerful enough to influence decision makers sitting in distant headquarters.
- The focus of Original Equipment Manufacturers (OEMs) was to build vehicles which due to their poor quality broke down forcing customers to return to the automobile dealership for repairs 'under warranty'.

NOTES

- In the marketing field, service marketing has emerged as a discipline of well-established research in the past two decades or so. In fact, service appears to be gaining ground over goods, as the fundamental for economic exchange.
- The crawling out stage witnessed the attempts made by authors to distinguish services from goods. This early phase was drawn out as the authors had an uphill struggle in introducing service as a fundamental provision for economic exchange.
- The concept of brand equity originated in order to measure the financial worth of this significant, yet intangible entity. Brand equity is the value and power of the brand that determines its worth.
- Awareness of the brand name among target customers is the first step in the equity building process. Awareness essentially means that customers know about the existence of the brand and can also recall what category the brand is in.
- Associations contribute to brand equity, as strong, positive associations induce brand purchases, besides generating good word-of-mouth publicity.
- Brand loyal customers form the bedrock of a company. Higher loyalty levels lead to a decrease in marketing expenditure as such customers act as positive advocates for the brand.
- Proprietary assets include patents, trademarks and channel relationships. These assets are valuable as they prevent competitors from attacking the company and prevent the erosion of competitive advantages and loyal customer base.
- Line extension leads to managerial focus on minor modifications, packaging changes and advertising rather than on real innovations. Cannibalization can also occur, i.e., the new brand variants gains sales at the expense of the established variants of the same brand.
- A new brand is the safest option for upscaling as it does not carry any baggage of associations with existing market segments. However, it is the riskiest and most expensive option as well.
- An effective international market research may facilitate marketers in identifying multiple options for introducing new products /services or options of entering new markets. International marketing decisions may prove less risky if they are supported by market research.
- All types of researches, whether for business, medical science, or the government are important for making decisions. There is also the risk that the results of these researches are wrong.

- Data and information from marketing information system should be available to both field and head office staff, when and where required, and in a format which aids decision-making.
- Branding is a major issue in product strategy. Developing a branded product needs a great deal of long-term investment spending especially for advertising, promotion and packing.

NOTES

14.9 KEY WORDS

- **Marketing Information System:** It refers to the systematic collection, analysis, interpretation, storage and dissemination of the market information, from both the internal and external sources, to the marketers on a regular, continuous basis.
- **Online Marketing Research:** It refers to the method of carrying out the marketing research process including the data collection over the internet.
- **Blockchain:** It is an online ledger which duplicates and distributes data across the entire network of computer systems on the blockchain. It uses safety measures like new ID verification.
- **Original Equipment Manufacturer (OEM):** It refers to an original equipment manufacturer is a company that produces parts and equipment that may be marketed by another manufacturer.

14.10 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Why is marketing research both a science and an art?
2. What is a Marketing Information System?
3. Write a short note on online marketing research.
4. List three ways of determining brand equity.
5. Why do companies engage in brand stretching?
6. What does international market research facilitate marketers to do?
7. List the various tests which can be conducted to test the branding policy soundness.

Long-Answer Questions

1. Discuss the reasons for the need of a marketing information system.
2. Elaborate upon some of the trends that can be seen in the field of marketing research.

3. Explain the three broad stages of the evolution of services marketing.
4. Discuss the steps involved in international market research.

NOTES

14.11 FURTHER READINGS

Saravanavel, P and S Sumathi. 2004. *Marketing Research and Consumer Behaviour*. New Delhi: Vikas Publishing House.

Kumar, Arun and N Meenakshi. 2016. *Marketing Management*. New Delhi: Vikas Publishing House.

Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing House.

Gupta, S.L. 2004. *Marketing Research*. New Delhi: Excel Books.

Malhotra, Naresh K. 2015. *Marketing Research*. New Delhi: Pearson Education.