Uttar Pradesh
Rajarshi Tandon Open University

MBA-3.44
Marketing Research

Block

1

MARKETING RESEARCH CONCEPTS AND DESIGN

UNIT 1

Marketing Research — Meaning and Importance 5

UNIT 2

Organisation of Marketing Research in India 13

UNIT 3

Research Design 23

Expert-Cum-Course Writers Committee

Prof. Rakesh Khurana
School of Management Studies
IGNOU, New Delhi

Prof. Abraham Koshy
Indian Institute of Management
Ahmedabad

Prof. Y.L. Grover
Management Consultant
New Delhi

Prof. Deepak Chawla
International Management Institute
New Delhi

Prof. Mukul Gupta
International Management Institute
New Delhi

Mrs. Neerja Wable
Indian Market Research Bureau
New Delhi

Shri Saurabh Khosla
MD, Tulika Advertising Agency
New Delhi

Prof. Sudas Roy
Indian Institute of Management
Calcutta

Dr. N. Kinra
Indian Institute of Management
Lucknow

Prof. A.P. Arora
Management Development Institute
Gurgaon

Mr. B. Narayanaswamy
Executive Director
Indica Research, New Delhi

Dr. P.S. Vishwanathan
Hindustan Computer Ltd.
New Delhi

Prof. Sharad Sarin
XLRI
Jamshedpur

Prof. Mohan Agarwal
XLRI
Jamshedpur

Mr. D.K. Bose
Media Director
R.K. Swamy, BBDO, New Delhi

Dr. (Mrs.) Madhulika Kaushik
Director
SOMS, IGNOU, New Delhi

Mr. Qasermuddin Haider
Asst. Regional Director
IGNOU, Patna

Prof. R.K. Grover
SOMS, IGNOU, New Delhi

Dr. Kamal Yadava
Course Coordinator
SOMS, IGNOU, New Delhi

Print Production : B. Natarajan, Copy Editor, SOMS, IGNOU

September, 2000 (Reprint)

© Indira Gandhi National Open University, 1998.

ISBN-81-7605-318-X

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the Indira Gandhi National Open University.

Further information on the Indira Gandhi National Open University courses may be obtained from the University's Office at Maidan Garhi, New Delhi-110 068.

Printed and published on behalf of Indira Gandhi National Open University, New Delhi, by Director, School of Management Studies.

Reproduced and reprinted with the permission of Indira Gandhi National Open University by G. S. Shukla Registrar, U.P.R.T.Open University, Allahabad (December 2011)
Printed by : Chandrakala Universl Put. Ltd. 42/7 Jawahar Lal Nehru Road Allahabad 211002

BLOCK 1 MARKETING RESEARCH CONCEPTS AND DESIGN

Block 1 on Marketing Research Concepts and Design consists of three units. Unit 1 discusses the meaning scope of the marketing research function and describes the stages in the marketing research process. Unit 2 on Organization of Marketing Research in India explains the reasons for growing importance of marketing research in India and the arrangements that India firms have for conducting marketing research. This unit also discusses the major applications of marketing research in India and the problems in doing marketing research in India. Unit 3 on Research Design discusses the utility of a research design in the research process, explains the different kind of research designs used in marketing research and describe the concept of causality.

UNIT 1 MARKETING RESEARCH — MEANING AND IMPORTANCE

Objectives

After going through this unit you should be able to

- Define Marketing Research
- Enumerate the critical factors in organising a good marketing research study
- Cite some of the possible application areas of marketing research
- Discuss the scope of the marketing redcarch furiction

Structure

1.1 Introduction

1.2 Marketing Research — Meaning and importance

1.3 Stages in the Marketing Research Process

1.4 Scope of the Marketing Research Function

1.5 Self-Assessment Questions

1.6 Further Readings

1.1 INTRODUCTION

The principal task of marketing management is to fulfil the aspirations of the consumers. It is thus imperative to understand what the consumers want; how they make the various choice decisions; or what are their sources of information and influence process etc. As such, marketing research is the function which provides the necessary information about the consumer to the marketer. In the proccss an organisation can idertify new opporunibes in the market evaluate and inoniwe marckeing actions and in general evolve better maeketing progamme to serve the interests of the consumer. Thus marketing research acts as the link between the consumer and the marketer.

1.2 MARKETING RESEARCH — MEANING AND IMPORTANCE

Operationally speaking, marketing research is defined as the objective and formal process of collecting information; analysing the results and communicating the findings and their implications in terms of marketing actions. However, some further comments about this definition are essential. Marketing rcsearch is a sysccmatlc collcetion and analysis of information that is ultimately used in evolving some marketing decisions. alll stages if a research study must be carried out in a logical manner. For instance, one should start wirh a concise statement of the issues to be investigated; indicate the information required to study those select problems; define the methods to be adopted to collect those data; specify the relevant technique to be employed for analysing the data; and finally state the reswarch findings and their specific implications for marketing decisions making.

Note that this definition indicates that marketing research should be conducted for specific issues. Secondly, it must ensure objectivity in every step. Finally, sndy fondiogs must help the manager in the decision making. In other words, marketingresearch must not be mere collection of statistical information. One must justify the choice of medodology of data collection and analysis and theresearchermust not be too much pro-occpied with techniques but instead cinvey the meaning of the results in the macketing language even

when some advanced or sophisticated tool is being used. Likewise, marketing manager(s) should also provide a clear, detailed scenario of the problems faced by the company before the marketing researchers (s). They must allow adequate time and budget for conducting the study. They must not use marketing research as a fire fighting device or to justify some preconceived action(s).

Marketing research is the function that likes the consumer with the organisation through information. It involves systematic and objective search for and analysis of information that can be used for evolving some marketing decisions.

Any research study must clearly state the issues being investigated. It must apply systematic and formal procedure in collection and analysis of information. It must communicate the study findings in a manner which could help in arriving at some marketing decisions.

A research study will fail to serve its purpose if marketing researcher merely collates some statistical facts or is pre-occupied with techniques or uses data of questionable validity or communicates the findings in too much vague or technical language.

Likewise, a research study will suffer if the marketing manager does not offer full perspective of the research problem; or allows inadequate time; or uses research as a fire-fighting device or does not really appreciate the value of research.

Problem must be clearly defined and reasons for undertaking the research from the point of view of marketing decision making should be explicitly justified.

In order to carry out effective research programme

1. Prepare a list of objectives to be examined
2. Avoid
   - — Vague terms of reference
   - — Trivial research projects
   - — Research where underlying purpose is unknown or with held.

3. Ensure concurrence about the terms of references (specially research objectives, plan of data collection, time and budget) among all concerned.

Activity 1

Talk to the marketing manager in your organisation or any other organisation that you are familiar with, about the following:

- Product decisions
- Promotion decisions
- Distribution decisions

How many of these decisions utilise formal marketing research? What additional research activity, in your view should have preceded them?

.....

.....

.....

.....

.....

.....

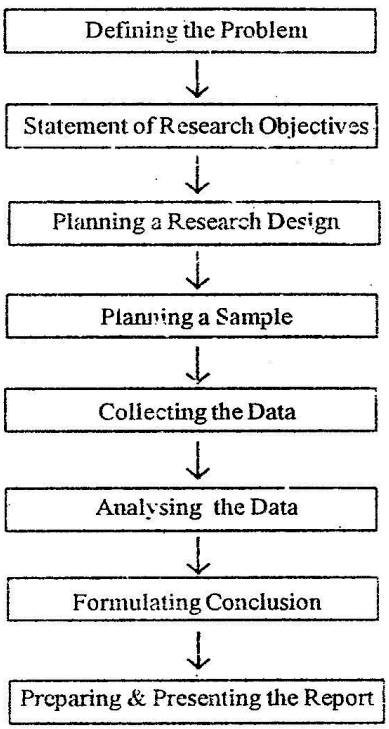
.....

.....

1.3 STAGES IN THE MARKETING RESEARCH PROCESS

Marketing research exercise may take many forms but systematic inquiry is feature common to all such forms. Being a systematic inquiry it requires careful planning of the orderly investigation process. Though it is an over simplification to assume that all research processes would necessarily follow a given sequence, marketing research often follows a generalised pattern which can be broken down and studied as sequential stages. Figure 1 gives the stages in the marketing research process.

Figure 1 : Stages in the Marketing Research Proces



As shown in Figure 1, the research process begins with the identification. The research task may clarify a problem or define an opportunity. The clearcut statement of problem may not be possible at the very outset of research process as often only the symptoms of the problem are apparent at that point. Thus the problem statement may be made only in general terms to be made specific later after some exploratory research has been done to clarify the problem situation.

Defining the Problem

Clear problem definition is of crucial importance in marketing research as in terms of both time and money research is a costly process. Careful attention to problem definition allows the researcher to set the proper research objectives which in turn facilitate relevant and economic data collection.

Problem definition in specific terms must precede the determination of the purpose of the research.

In order to define the problem more precisely, some sort of exploratory research may also be undertaken. The methods popularly in use are survey of secondary data, experience survey or pilot studies.

Statement of Research Objectives

After clarifying and identifying the research problem with or without exploratory research, the researcher must make a formal statement of research objectives. Research objectives may be state in qualitative or quantitative terms and expressed as research question statements or hypothesis. For example, the research objective "To find out the extent to which the sales promotion programmes affected sales" is a research objective expressed as a statement. A hypothesis on the other hand is a statement that can be refuted or supported by empirical findings. The same research objective could be stated as : "To test the hypothesis that sales are positively affected by the sales promotion programme undertaken this summer." Example of another hypothesis may be "Concentrating advertising efforts in monthly waves (rather than advertising continuously) would cause an increase in sales and profits." Once the objectives or the hypothesis are developed the researcher is ready to choose the research design.

Planning the Research Design

Once the research problem has been defined and the objectives decided, the research design must be developed. A research design is a master plan specifying the procedure for collecting and analysing the needed information. It represents framework for the research plan of action. The objectives of the study discussed in the preceding step are included in the research design to ensure that data collected are relevant to the objectives. The researcher must, at this stage, also determine the type of sources of information needed. the data collection methods (surveys or interviews for example), the sampling methodology and the timing and possible costs of research. The design chosen may be from exploratory, descriptive, quasi-experimental or experimental design categories which again include a number of alternative methods. You will study more about specific types of research designs and their appropriateness for particular kinds of research problem in unit 3 entitled 'research design'.

Planning the Sample

Although the sample plan is included in the research design, the actual sampling is a separate and important stage in the research process. Sampling involves procedures that use a small number of items or parts of the population to make conclusion regarding the whole population. The first sampling question that needs to be asked is who is to be sampled. which follow from what is the target population. Defining the population may not be as simple as it seems. For example, if you are interested in finding the association between savings and loans, you may survey the people who already have accounts and the selected sample will not represent potential customers.

The next important issue is regarding sample size. How large or how small should a sample be? Generally speaking, larger samples give more reliable information than smaller ones but if probability sampling is used, a small proportion of the population may give a reliable measure of the universe. Unit 5 on 'sampling' gives a detailed account of determination of sample size as well as different sampling methods used. The researcher is also required to know how to select the various unit to make up the sample. There are two basic classes of sampling methods-probabilistic, and non-probabilistic, both of which have been discussed at length in unit 5.

Data Collection

The data collection process follows the formulation of research design including the sampling plan. Data which can be secondary or primary, can be collected using variety of tools. These tools are classified into two broad categories, the observation methods and the communication methods, all of which have their inherent advantages and disadvantages. All these methods and their appropriateness for different situations have been discussed in detail in unit 4 on 'data collection'

Data Processing and Analysis

Once the data has been collected it has to be converted to a format thay woll suggest amswe to the problem identified in the first step. Data processing begins with the editing of data and Coding Editing invoived dpecting the data collccton forms for oamissoon icgioms a comsiscnes of classificnrion

Before tabulation, responses need to be classified into meaningful categories. The rules for categorising, recording and transferring the data to data storage media are called codes. The coding process facilitates the manual or computer tabulation. If computer analysis is being used, the data can be key-punched and verified.

Analysis represents the application of logic to the understanding of data collected about the subject. In its simplest forms, analysis may involve determination of consistent patterns and summerising of appropriate details. The appropriate analytical techniques chosen would depend upon informational requirements of the problem, characteristics of the research designs and the nature of the data gathered. The statistical analysis may range from simple univariate analysis to very complex multivariate analysis. You will study univariate, bivariate and multivariate analysis and their applications in marketing problem in last three blocks of this course.

Formulating Conclusion, Preparing & Presenting the Report

The final sate in the research process is that of interpreting the information and drawing conclusions for use in managerial decisions. The research report should effectively communicate the research findings and need not necessarily include complicated statements about the technical aspect of the study and research methods. Often the management is not interested in details of research design and statistical analysis but in the concerete findings of the research. If executives are to act on these findings they must be convinced of the value of the findings. Researchers, therefore, must make the presentation technically accurate, understandable and useful.

Frequently, the researchers are required to make both an oral and a written presentation. Since each project is different, the presentation in each case requires originality. However, the better the earlier steps in the research process have been executed, the more likely it is that a good presentation would result. While the oral presentation depends a great deal on the personal style of the presenter and the management expectation, the written report in order to be effective needs to include the following details:

I) Title page
II) Table of contents
III) Introduction
IV) Statement of objectives
V) Methodology
   A) Research design
   B) Data collection methods
   C) Sampling
   D) Field work
   E) Analysis and interpretation
VI) Limitations
VII) Findings
VIII) Conclusion and recommendations
IX) Appendix
   A) Copies of forms used
   B) Tables not included in finding
   C) Bibliography, if relevant

Activity 2

With respect to a marketing research project **undertaken in your organisation or any other** organisation outline the steps followed in the research.

.....

.....

1.4 SCOPE OF THE MARKETING RESEARCH FUNCTION

This section will discuss the possible applications of marketing research in the corporate environment. Bulk of research is done to measure consumer wants and needs. Also, marketing research is carried out to assess the impact of past marketing actions. Some research is done to understand the competitive, technological, social, economic, cultural, political or legal environments of the market.

Another way of looking at the function of marketing research is to look at the particular decision area where research results are used.

I) **Sales Analysis**
Much research is done in the following areas which are broadly referred as sales analysis.

- Measurement of market potential/demand projection;
- Determination of market characteristics;
- Market share estimation;
- Studies of business trends.

In fact, some of the more detailed studies to be carried out under the broad ambit of sales analysis could be as follows.

- The types of consumers that constitute the potential market;
- The size and location of the market;
- The growth and concentration of the market over certain period of time;
- The competitive picture for the product;
- The major strategies of leading competitors with respect to price, offerings distribution etc.
- The purchase habits of key market segments;
- What is the pattern of pre-purchase deliberations made by the consumers?
- Who are involved in the decision making?
- What are the role of different members in the decision making?
- How does the product fit into the consumer's life styles? (Operation if it is an industrial product);
- Do consumers prefer to buy scme particuinr bramcis ? (i.e. Assess the degree of brand loyalty

The above list is not exhaostive Here researcrt is basically done with a view to know consumers' motivation, attitude, cognition and perceptions etc. Thus information will be collected in a manner sothaat they have some implications for various marketing decisions.

II) **Sales Methods and Policies**
Marketing research studies are also condocted with a view to evalnatc the effectivness of present distribution system. Such studies are used in establishing or revising sales territories. They are aldo helpfut in establishmcm of sales quotas, design of territory boundary, compcnsation to sales force ohysical distrjbulion and distribulion cost analysis, etc. Marketing research is also done ti assess the cofectiv eness of different promotional activioes such as premims dealds coupons sampling rtc

III) Product Management
Every marketer tries to fotmally cr informally utilize information to manag the

existing and new products. It examines market feedback about competitive product offerings. Also, some companies make use of marketing research to form market segments through choice of alternative bases. Companies also carry out different research studies to assess consumer feedback to new products and their likely potential. Of late, in India many consumer products have been launched after making rigorous amount of research. Moreover, researches have enabled to diagonise how consumers perceive various brands of a product. Such studies have enabled the companies to position their brands.

Marketing research studies have been conducted to monitor the performance of the test brand (in terms of trial and repeat purchase) launched in select shops in the market. Studies are popularly known as simulated test marketing (STM) models.

Pricing studies, packaging research, design or physical characteristics, have also been sometimes conducted.

IV) **Advertising Research**

Media research: Three National Readership Surveys (NRS) have so far been conducted in India. These studies have basically estimated the readership of leading newspapers. The last NRS has also assessed, qualitatively, readers feedback on the editorial content

Moreover, some marketing research have evaluated the relative effectiveness of different media in specific product fields, and in context of achieving specific tasks such as creating brand awareness or a particular product benefit.

Copy research: Advertising agencies have been regularly engaged in this activitywhere they test out alternative copy designs by obtaining the feedback from te consumers.

Studies of advertisement effectiveness — Advertising agencies regularly make use of marketing research studies to assess and monitor effectiveness of different advertising compaigns.

V) **Corporate Research**

Large scale corporate image studies among different target publics — They involve an assessment of knowledge about company activities, association of company with sponsored activities and company perceptions on specific dimensions. These types of corporate image studies are done periodically to monitor any change in image over time among different publics.

Social values research: Knowledge, attitude and practices on family planning, anti-dowry, smoking, drinking etc.

Political studies: In recent times marketing studies have been conducted to ascertain the public opinion about the election results.

Customer service studies: Many banks and large industrial houses have resorted to marketing research to know the consumers' changing need for service and possible grievances about existing operations.

VI) **Syndicated Research**

Several research agencies collect and tabulate marketing information on a continuing basis. Reports are sent periodically (Weekly, monthly or quarterly) to clients who are paid subscribers. Such services are found specially useful in the spheres of movement of consumer goods through retail outlets (ORG Retail Audit), incidence of disease and use of branded drugs (MARG- prescription audit), Television Program viewing (the Television Rating Points), Newspaper & Magazine readership (NRS - discussed earlier under media research), assessment of market potential of a city with population one lakh and above (Thompson Indices), study of nation's attitudes and psychographics (PSNAP and IMRB's life style research on the cigarette market).

1.5 SELF-ASSESSMENT QUESTIONS

1) What is marketing research?
2) Briefly comment on the definition of marketing research.
3) Indicate whether marketing research is relevant to each of the following organization and if so, how each might benefit fromit.
   a) Your company
   b) Central Ministry in the Government of India
   c) A retail shop
   d) A bank
   e) A service business
4) What are the major weaknesses of marketing research?
5) Suggest some of the precautions to be adopted in the problem definition stage.
6) Suggest some possible areas of application of marketing research.

1.6 FURTHER READINGS

Marketing Research: Text and Cases — H.W. Boyd. Ralph Westfall and S.F. Stasch. Homewood, Lllinois, Richard D. Irwin Inc.,

Basic Marketing Research — Gilbert A. Churchill (Jr.) Hinsdale, Illinois, Dryden Press,

Handbook of Marketing Researh — Rober Ferber (Ed.) New York, McGraw Hill.

Research for Marketing Decisions — Paul E. Green and Donald S. Tull. Englewood cliffs N.J. Prentice Hall. Inc.,

Marketing Research: Applications and Problems — Arun K. Jain. Christian Pinson and Brian T. Rachford. New York. Wiley,.

Marketing Research: An Applied Approach — Thomas C. Kinner and James R. Taylor, Singapore, McGraw Hill.

Marketing Research: Text, Applications and Case Studies — Ramanuj Majumdar, New Delhi, Wiley, 1990.

Marketing Research: Information Systems and Decision Making — Bertrem Schoner and K.P. Uhl, New York, Wiley.

Marketing Research: Meaning, Measurements and Methods — D.S. Tull and D.I. Hawkings New York: Macmillan..

Marketing Research: Fundamentals and Dynamics — Cerald Zaltman and P.S. Berger. Hinsdale, Dryden Press.

UNIT 2 ORGANISATION OF MARKETING RESEARCH IN INDIA

Objectives

After studying this unit, you should be able to:

- know as to why marketing research is becoming important in India,
- describe arrangements that the Indian firms have for conducting marketing researches,
- enumerate major applications of marketing research, and
- understand problems faced in doing marketing research in India.

Structure

2.1 Introduction
2.2 Growing Importance of Marketing Research in India
2.3 Expenditure on Marketing Research in India and Major Users
2.4 Ways of Conducting Marketing Research in India
2.5 Increased Demand for Outside Research Services
2.6 Common Applications of Marketing Research
2.7 Problems in Conducting Marketing Research in India
2.8 Summary
2.9 Key Words
2.10 Self-Assessment Questions
2.11 Further Readings

2.1 INTRODUCTION

You have learnt meaning, nature and importance of marketing research in Unit 1. You have also gained familiarity with steps involved in conducting marketing research. In this unit, we shall discuss the role and organization of marketing research in India. More specifically, we shall explore as to why marketing research is becoming important in India and what organizational arrangements firms have for conducting researches. Major users and uses of marketing research in the country shall be identified . The unit shall also examine the problems which the researchers face while doing researches in India.

2.2 GROWING IMPORTANCE OF MARKETING RESEARCH IN INDIA

Use of marketing research in India has been considerably low in the past. Only a few large firms, especially those having linkages with foreign firms, were making use of marketing research during the fifties and sixties. Even among such firms, only a select ones were conducting researches on a regular basis and/or having a marketing research department within the organization. A number of reasons can be put forward to explain this phenomenon. Highly protected nature of the Indian market during the fifties and sixties is amongst the most important reasons. A highly restrictive industrial licensing policy coupled with a ban on capacity expansion led to constant shortages of goods and services in the country. With demand always exceeding supply and a virtual absence of competition in the market, the firms were able to easily push to the consumers all those

products they produced. The focus of the firms at that time was primarily on procuring the licenses and raw materials. They hardly had any concern for the product improvements or satisfaction of their customers' needs and wants. In a way, a sort of sellers' market prevailed at that time, and availability of an assured market to the producers did not provide any incentive to them to make use of marketing research for becoming more efficient, innovative and competitive in the market place.

Moreover, the firms during those times operated on small scales and that too within the peripheries of the local markets. Since they were very close to their customers and had personal knowledge about them, they did not feel the need for conducting marketing researches. Even a few of the firms that were desirous of making use of marketing research could not do so because of the paucity of funds. Lack of professional approach to management and decision making was another important reason for the neglect of marketing research among the business firms in India.

Of late, however, the situation has undergone considerable changes. With successive deregulation of the economy in the ensuing period and abolition of the archaic industrial licensing system and other regulations especially in the nineties, the competition has hot up in the market. Many a new firms - both the Indian and foreign ones - have made a foray into the market. Scalewise also, firms have expanded their businesses and they today operate in the local as well as national and international markets. Due to fast changing technological environment, product life cycles have been shortened and many a new substitute have emerged. New methods of communications and distribution channels have radically transformed the Indian marketing scene. In many product categories, sellers' markets have paved way to the buyers' markets. In the wake of these changes and growing customer sophistication, it is no longer possible for the firms to continue to base their marketing decisions solely upon gut feelings and past experiences. Need for timely, accurate and relevant information about the markets and buying behavior of the target market customers is fast becoming *sine qua non* for improved decision making.

2.3 EXPENDITURE ON MARKETING RESEARCH IN INDIA AND MAJOR USERS

Though no precise information is available; the guesstimate is that Indian firms do not spend much money on marketing research. Consulting and Research Enterprise (CORE), Hyderabad conducted a survey of the Indian firms in 1986 to ascertain their expenditure on marketing research and attitudes toward it. About 500 questionnaires were mailed out to the firms selling industrial goods, consumer products and services, out of which 130 responded. Despite an overall favorable attitude towards marketing research, the study found most of the firms spending low money on it. About 40 per cent of the firms were found spending less than Rs. 1 lakh. Expenditure by the other 46 per cent of the firms ranged between Rs. 1 lakh and Rs. 10 lakhs. Only about 14 per cent firms we spending more than Rs. 10 lakhs, with highest figures being Rs. 1 crore (see Table 1).

CORE study also attempted an analysis of the marketing research expenditure in relation to sales, and found that while 28 per cent of the surveyed firms were spending 0.01 per cent or even less of their sales volume on marketing research during the year 1985-86, 63 per cent firms had their expenditure on marketing research in the range of 0.02 per cent to 0.4 per cent of sales. Though the rest of the 9 per cent of the surveyed firms were spending more than 0.4 per cent on marketing research, it was only the firms producing consumer goods which had expenditure on marketing research in excess of 1 per cent of their sales.

Table 1
Expenditure on Marketing Research* by Indian Firms : 1985-86

Expenditure	Percentage of respondents
Amount	
Less than Rs. 1 lakh	40.0
Rs. 1 lakh to Rs. 10 lakhs	46.0
More than Rs. 10 lakhs	14.0
	100.0
Marketing Research Expenditure as a Percentage of Sales	
Up to 0.01	28.0
0.02 - 0.09	34.0
0.10 - 0.40	29.0
More than 0.40	9.0
	100.0

Source: Busiess India November 17-30-1986
Note: Inclusive of expenditure on in house as well as outside rescarch services

In the past, only big firms used to be major users of the marketing research. But of late, medium and small size firms have also started showing interest in marketing research. Little wonder that even the marketing research agencies have started turning to them for securing research business from them. The number of the firms using marketing research on a regular basis is also on an increase.

Over time, a significant decline has occurred in the relative importance of the multinational firms. Two decades ago, multinational corporations were the major users of marketing research, and their share was as high as 67 per cent of the total commercially available research in India. But their share has now started declining, and is expected to fall to a level of 45 per cent by the end of this century. In contrast, the Indian firms are fast emerging as the major users of the marketing research in the country

Another noteworthy change is the increasing use of marketing research by the industrial goods firms. Though the consumer goods firms continue to dominate the scene even at present, the share of industrial goods firms is likely to rise in future with increased competition in the industrial goods market.

2.4 WAYS OF CONDUCTING MARKETING RESEARCH IN INDIA

Broadty spcaking a firm one get the rescarchcy conducted in two wayay cither through its own staff or by hiring the services of outside agencies. Some firms make use of both the alternatives.

So far as the in-house research is concerned, it can be conducted by the firm's sales and marketing staff of or else they can have specialized staff or department for conducting marketing researches. Since the employment of specialized research staff or setting up of a separate research department entails huge expenditure, only large firms requiring researches on a frequent basis go in for this type of organizational arrangement. Small firms and the organizations not so frequently requiring researches prefer using their sales or marketing staff for undertaking research studies. As and when the need arises, some of the sales and marketing persons are picked up and assigned the rescarch tesks sonee these pcrsons gchcralfy do not have spccialized knowlcdge of

marketing research and, moreover, perform side by side the routine sales and marketing activities, quality of the studies conducted by such persons remains under doubt.

The other important method of collecting information is the use of outside agencies or experts such as marketing research organizations, advertising agencies and consultants. Seeing the increased demand for research services, many a marketing research organization have come up in the country and help business firms through provision of customized as well as syndicated research services. When the research is tailor-made to meet the information needs of one particular client, it is known as customized research. Since the benefit of such a research accrues to a single firm, that single firm alone bears the entire research cost. Syndicated research, on the other hand, involves collection of a common pool of data for meeting the information needs of a number of firms. As the results of such a study are shared among several firms, it costs pretty less to a single firm. Quite often, these studies are undertaken on a regular basis by the research agencies and the information so collected is made available to the firms on a subscription basis.

Table 2 contains the information regarding usage of different modes of conducting researches. It is evident that the majority of the Indian firms (about 70 per cent) depend upon their sales staff for doing researches. Use of outside marketing research agencies comes next in order of importance, with about one-half of the firms employing them. In-house marketing research staff exists in the case of 44 percent of the firms, especially among the large size firms and those engaged in marketing of industrial products and services. About one-third each of the Indian firms get their marketing research studies done through advertising agencies and consultants.

Syndicated researches are not yet common among the Indian firms. Despite their low costs to the clients and the advantage of such information in many cases being available at regular intervals, only about one-fifth of the firms make use of the syndicated research services.

Table 2
Organizational Arrangements for Conducting Marketing Research in India

Arrangement	Percentage of firms*	Nature of the firms preferring the arrangement
Firms' sales staff	69	Firms of all sizes, industrial goods and services marketers
Marketing research agencies	52	Large firms, consumer goods marketers
In-house marketing researchers	44	Large firms, industrial goods and services firms
Advertising agencies	35	Smaller firms
Consultants	34	
Syndicated research services	22	Large firms and consumer goods firms

Source: Adapted from Business India, November 17-30, 1986.
Note: The percentage figure will not add up to 100 as many firms reported using more than one arrangement.

Table 2 also provides information about the nature of the firms preferring a particular arrangement. It can be observed that practice of using sales staff for carrying out

researches exists among the firms of all sizes. Use of in-house marketing research staff, outside research agencies and marketing research services is more often prevalent among the larger rather than medium and small size firms. Important differences also exist among firms selling different types of products. While industrial goods and services firms show a greater reliance on their sales persons and in-house marketing research staff, consumer goods firms depict a preference for marketing research agencies and syndicated research services to meet their information needs.

Activity 1

How is marketing research activity organized in your firm? List the arrangements that **your firm has for conducting** researches both within and outside the firm. (Alternatively **you may approach some other** organization and complete this activity in respect of that **organization**).

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

2.5 INCREASED DEMAND FOR OUTSIDE RESEARCH SERVICES

The demand for outside marketing research services has phenomenally increased during the eighties and nineties. This is evident from the increase in turnover of the marketing **research agencies during this period. The value of marketing research conducted by the** marketing research agencies was just Rs. 1 crore during the seventies. It went up to a figure of Rs. 3 crore in 1983, and further up to Rs. 10.06 crore in 1988-89. Currently, the turnover of the research agencies is expected to be somewhere around Rs. 20 to Rs. 25 crore.

Seeing growth in demand for different kinds of outside marketing research services, a number of marketing research agencies and other organizations have mushroomed in the country to meet the information needs of the business firms. A list of some of these organizations is provided in Table 3.

Table 3

Major Organizations Offering Marketing Research Services in India[1]

Operation Research Group (ORG)[2]
Indian Marketing Research Bureau (IMRB)
Marketing and Research Group (MARG)[2]
Marketing Operations Design and Enquiry Service (MODES)
Pathfinders
Advertising Consultants (India) Pvt. Ltd.
Concept Marketing and Advisory Pvt. Ltd.

Marketing and Business Associates (MBA)

Research Consultants and Marketing Services Pvt. Ltd.

Consulting and Research Enterprise (CORE)

Marketing Research and Data Processing Consultancy

Tata Consultancy Services

National Council of Applied Economic Research (NCAER)

Indian Institute of Public Opinion

Indian Institute of Foreign Trade (IIFT)

Feedback Marketing Services

Research Consultancy Group

C. Marck (India) Ltd.

Quantum Market research Pvt. Ltd.

Research International India

Indica Research

Source: Adapted from Sanjay K. Jain, "Marketiag Research in India : A Perspective" *Management & Change*, Volume 1, No. 1, April 1997, p.82.

Notes

1. It is only an illustrative rather than exhaustive list of the organizations rendering marketing research services in India.
2. ORG and MARG have recently joined together and the new outfit is known as ORG-MARG.

Activity 2

Compile a list of the organizations that provide marketing research services in your area. Also approach any two of them and find out what type of research services they provide to the clients.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

A break-up of the firms using services of marketing research agencies is provided in Table 4. Manufacturing firms have been the single largest users of such research services, with a share of 72 per cent of the total commissioned marketing research in India. Though the advertising agencies are themselves important research services providers, they also employ marketing research agencies for getting researches conducted for them. Share of the firms using syndicated research services has been just 6 per cent in 1989. Media owners, services industries, government departments/organizations, and research/consultancy organizations are yet to emerge as important users of the commissioned marketing research in India.

Table 4

Clientwise Break-up of the Marketing Research Commissioned in India during 1988-89

Client	Value (Rs. in crore)	Percentage share
Manufacturing companies	7.21	72
Advertising agencies	0.85	9
Syndicated/subscriber studies	0.63	6
Media owners	0.51	5
Service industries	0.32	3
Government department /agencies	0.21	2
Research/consultancy organizations	0.21	2
Others	0.12	1
Total	10.06	100

Source: Sanjay K. Jain, "Marketing-Research Services Marketing" in Ravi Shankar, ed., *Services Marketing : The Indian Experience*, South Asia Publication, Delhi, 1997, p.397.

2.6 COMMON APPLICATIONS OF MARKETING RESEARCH

The most common application areas of marketing research in India are listed in Table 5. New product decisions constitute the most frequent usage of marketing research. The other important application areas include: estimation of market share, collection of competitive information, demand estimation, product modification decisions, measurement of customer satisfaction, product positioning, diversification and market segmentation decisions. Only about one-third of the firms reported the use of marketing research for developing advertising theme/message and arriving at pricing and customer service decisions.

Application of marketing research for product elimination and evaluation of advertising effectiveness is prevalent only among 20 per cent firms. Furthermore, only 9 per cent firms reported the incidence of marketing research for channel modification decisions.

Table 5

Common Applications of Marketing Research in India

Areas in which marketing research is frequently/always used	Percentage of respondents*
New product decisions	58
Estimating market share	49
Gathering competitive information	47
Demand estimation	46
Product modification decisions	44
Measuring consumer satisfaction	43
Product positioning decisions	42
Diversification decisions	40
Market segmentation decisions	38
Advertising theme/message decisions	31
Pricing decisions	30
Customer service decisions	27
Product elimination decisions	21
Evaluating advertising effectiveness	20
Channel modification decisions	9

Source: **Business India, November 07-30-1986**

Note: * **Percentage fogures will not add up 100 as many firms reported multiple uses of marketing research**

Activity 3

For what purposes has your organization in the last two years conducted or commissioned marketing researches? List the major applications. (Alternatively you mey approach another organization and complete this activity in respect of that organization).

2.7 PROBLEMS IN CONDUCTING MARKETING RESEARCH IN INDIA

Due to country's vast size , heterogeneous population and infrastructural and attitudinal problems, it is not easy to conduct marketing research in India. Some of the major problems that you as a researcher would face in conducting marketing research in the country are discussed below.

India's large and heterogeneous population comes in a big way in conducting marketing research. Being a big and diverse country, national surveys require India to be divided into several hundred districts and interviewing several thousands of people. This calls for enormous time and money and a big fleet of field workers - well beyond the capacity of any small or medium size company.

Cultural diversity and linguistic nuisances further compound the researcher's problem. More than 14 languages are spoken in the country, with dialects exceeding 1,400 in number. Any major survey in the country requires translation of the questionnaire in a minimum of five to six languages. Many a time, strict translation of certain technical words or phrases is not possible, thus giving rise to the problem of non- comparability of data across the regions.

Accessibility to people living in the hinterland of the country is another big problem. Only very few people own telephone. Postal system is also not up to the mark. Because of low literacy level, mail interviews are of limited application. Personal interviews seem to be the only viable alternative, but even these are beset with transport problems and lack of trained staff in the small towns and rural areas.

Secondary data available in the country also suffer on account of poor coverage and redundancy of information. Data are at all not available for many a variable of interest to the researchers. Though census is conducted after every ten years in the country, it is after a considerable lapse of time that the full results are released. Even the trade and industry associations lack complete and uptodate lists of the manufacturers and traders. The industry and firms' production and sales figures are also not complete, uptodate and reliable.

Use of random and other elaborate sampling techniques presuppose the existence of suitable sampling frames (i.e., list of the target market population from which the samples are drawn). Non-availability of such lists in the country complicates the research tasks and forces the researcher to use non-probability sampling methods, thus adversely affecting the reliability and validity of the survey results.

Attitudinal problems also restrict the usage of marketing research in India. The study by Consulting and Research Enterprise (CORE) group, for instance, found that only two-third of the executives of the surveyed firms had the opinion that marketing research findings represent the real world, and marketing research data are reliable enough to be of use in decision making. In response to the question whether costs incurred on marketing research are low relative to the benefits that accrue from it, about 58 per cent firms indicated disagreement, implying low utility of marketing research. Further only 57 per cent of the executives refuted the statement that "gut feeling is more important than marketing research". Rest were either ambivalent or in agreement with the statement.

The managers also appeared quite concerned with the time involved in completing the marketing research studies. Only 42 per cent of the respondents did not feel that " market research often takes too long to be of any real use". In terms of quality and sophistication too, marketing research in India in the opinion of many executives is far below the expectations.

Not only the business firms, but advertising agencies also do not hold favorable attitudes towards marketing research. Advertising executives view it as a hindrance to their creative work and hence do not like spending much money on it. In 1988, about Rs. 1078 crores were spent on advertising in the country. Had even 1 per cent of this amount been spent on marketing research, the expenditure on advertising research alone would have been Rs. 10.78 crore. But this is not true as the total expenditure on commissioned marketing research itself in India was just of the order of Rs. 10.06 crore in 1988-89.

2.8 SUMMARY

Because of excessive regulations and lack of competition in the market in the past, not much need was felt for marketing research in India. However, with the dismantling of the licensing system and other regulatory measures during the last two decades or so, and increasing competition in the market, fast changing technology and growing consumer sophistication, it is no longer possible for the marketing managers to solely make decisions on the bias of their gut feeling and past experiences. They have come to increasingly rely upon the information collected through the process of marketing research. The Indian firms make use of both the in-house staff and outside research organizations, advertising agencies and consultants for getting researches conducted. Seeing the growth in demand for outside research services, a large number of marketing research organizations have come up in the country to provide both the customized and syndicated research services.

In view of the country's large size, heterogeneous population, language diversity, infrastructural bottlenecks and lack of secondary data, it becomes a uphill task for the researcher to carry out researches in India. Unfavorable attitudes of the managers also come in a big way in the widespread acceptance of marketing research in India.

2.9 KEY WORDS

In-house research : Research conducted by the staff within the firm.

Outside research services : Hiring services of the persons, organizations and agencies outside the firm for conducting researches. The outside organizations include marketing research agencies, advertising agencies and consultants.

Marketing research agency/organization : It is a firm or an organization which specializes in the area of marketing research and earns revenue by conducting researches for its clients.

Customized research : A type of research conducted by the marketing research agency for its client. The whole research project is designed to collect the information as required a particular client.

Syndicated research : A type of research in which the research agency collects a pool of data for serving the information needs of a number of clients. Since the information so collected is sold to a number of clients, it costs much less to a single firm purchasing such information.

2.10 SELF-ASSESSMENT QUESTIONS

1. What are the major reasons for growing importance of Marketing Research in India?
2. What are the different ways of conducting marketing research? Describe the scope for outsourcing of marketing research services in India?
3. Suggest some possible areas where marketing research would find increasing application in India?
4. Briefly comment on the problems faced by researchers in conducting marketing research in India?

2.11 FURTHER READINGS

Beri, G.C., 1993, *Marketing Research*, 2nd Edition, Tata McGraw-Hill Publishing Company Ltd. : New Delhi, pp. 88-95. These pages provide an overview of the syndicated research services available in the country.

Jain, Sanjay K., 1997, " Marketing Research Services Marketing " in Ravi Shankar (ed.), *Services Marketing: The Indian Experience*, South Asia Publications ; Delhi, pp. 392-415.

Jain, Sanjay K., 1997, " Marketing Research in India : A Perspective", *Management & Change*, April, pp. 71-100.

Rao, Sushila, 1986, "Marketing Research in India : The State of the Art", *Business India*, November 17-30, pp. 111-115.

UNIT 3 RESEARCH DESIGN

Objectives

After studying this unit you should be able to:

- Explain the concept and importance of research design in the context of marketing research problems
- Discuss the causality factor in the context of research design
- Describe the various kinds of research designs and their applications for different research situations.

Structure

3.1 Introduction
3.2 Research Design — Meaning and Importance
3.3 Causality: The Basis of Classification of Various Types of Research Designs
3.4 Exploratory Research Design
3.5 Descriptive Research Design
3.6 Factors Influencing Experimental Validity
3.7 Quasi-Experimental Designs
3.8 Experimental Designs
3.9 Summary
3.10 Self-Assessment Questions
3.11 Further Readings

3.1 INTRODUCTION

In the last two units you studied about the nature and importance of marketing research. In this unit you will learn about the first phase in planning the research project, which is formulating a research design. Based on causality, research designs have been divided into four categories of exploratory, descriptive, quasi-experimental and experimental designs. While exploratory and descriptive studies provide information on events and attributes from which inferences can be drawn, they can only offer tenuous conclusions. In order to draw conclusions it is important to formulate quasi-experimental or experimental designs. This units describes different types of designs within the four categories mentioned above as well as the marketing situations where these designs could be most appropriately applied.

3.2 RESEARCH DESIGN—MEANING AND IMPORTANCE

The research design is a comprehensive master plan of the research study to be undertaken, giving a general statement of the methods to be used. The function of a research design is to ensure that requisite data in accordance with the problem at hand is collected accurately and economically. Simply stated, it is the framework, a blueprint for the research study which guides the collection and analysis of data. The research design, depending upon the needs of the researcher may be a very detailed statement or only furnish the minimum information required for planning the research project. To be effective, a research design should furnish at least the following details.

a) A statement of objectives of the study or the research output.

b) A statement of the data inputs required on the basis of which the research problem is to be solved.

c) The methods of analysis which shall be used to treat and analyse the data inputs

Let us try to understand the elements through an example:

A newly opened supermarket sells a broad line of merchandise ranging from provision to household appliances and kitchenware. The general manager (sales) believes that the total profits of the supermarket can be enhanced by getting people to buy in larger quantities which could be achieved by offering attractive cash discount on bigger purchases. As the other executives are doubtful about this. The hypothesis can be tested by carrying out a marketing research exercise.

1) The objective is to calculate the margin earned on sales when this discount is offered and compare it with the margin when discount is not being offered.

2) The data to be collected over a period of time
   a) Sales in rupees to a selected sample of customers during the period when the discount is offered.
   b) Sales in rupees to the same customer when the discount is not being offered.
   c) The average order size in the two periods.
   d) The average margins earned during the two periods.
   e) The cost of promotional inputs regarding the discount.

3) The analysis of the data will be
   a) Sales in rupees in period I minus those in period II.
   b) Subtract cost of incentives.
   c) Also subtract cost of promotional inputs.

The importance of research design lies in the fact that it makes a statement of what is to be done in order to achieve the research objectives and how it is to be done. It is an expression of what is expected of the research exercise in terms of results and the analytical input needed to convert data into research findings.

The research design furnishes a clear idea as to the activities that would need to be undertaken in order to achieve the research objective. It is therefore, helpful if the research design after being finalised is put in writing to enable the researcher to have a frame of reference and prevent the study from deviating.

At the stage of analysis and interpretation also, the research design helps in providing direction to the computation and interpretation process to arrive at solution and recommendations. This is however not to suggest that a design is a rigid straitjacket to which the study must always conform. The research design only represents an expectation of likely results but as the study proceeds, many unexpected results may come forth which may necessitate framing of new hypothesis or at least modifying some. The research design is only a guiding and not a limiting framework for research study.

3.3 CAUSALITY: THE BASIS OF CLASSIFICATION OF VARIOUS TYPES OF RESEARCH DESIGNS

There are four types of research designs which are broadly classified as:

1) Exploratory Designs
2) Descriptive Designs
3) Quasi-Experimental Designs
4) Experimental Designs

These designs will be discussed in detail very shortly. However, one point may be noted that the basis of classification of various types of research designs is causality, which is the subject matter of discussion of this section.

Let us start our discussion with an example. Suppose the sales manager of a company manufacturing table fans carried out a training programme for its salesmen operating in a state. Three months after the training programme, it was observed that there was an

increase in sales in that state by 40 per cent. The sales manager claimed that the training programme was effective and therefore, the salesmen operating in other states should also undergo the same training programme. We can see that the sales manager is inferring that the training programme has resulted in a sales increase of 40 per cent. That is to say that training programme is the cause of increased sales.

The sales manager is referring to sales training as a causal variable and the sales increase as the effect variable. We may now ask a question whether this statement is valid? Has the sales training really resulted in increase in sales? The possible answer is that we cannot say with certainty that the sales training has caused the increase in sales. There could have been other factors which might have been responsible for the increased sales. Increase in sale might have been caused by decrease in the price of the table fans, a strike at a competitor's factory, increased product penetration in the distribution channel, weather conditions, etc. Therefore, it is very essential that the sales manager should know that conditions under which proper causal statements can be made. To be able to make any causal statements, one should ensure whether the following three conditions are met :

i) We must have very strong evidence to say that there exists a strong association between an action (causal variable) and the ultimate outcome (effect variable).

ii) The other condition for the causal relationship is that the action (causal factor) must precede the observed outcome (effect variable).

iii) We must have strong evidence to say that there were no other possible factors (causal factors) which could have resulted in the observed outcome.

The first condition is that of concomitant variation. Concomitant variation is the extent to which a cause, X, and effect Y occur or vary together. In our example, the sale training programme and increase in sales would need to occur together. To be able for us to conclude causality, the condition of concomitant variation between variables in question must exist. However, it may be worth mentioning that a strong association between two variables does not constitute a proof of a causal relationship. It is quite likely that the strong association between two variables may be as a result of random variation or both the variables may be influenced by an extraneous variable.

The second condition for causal relationship is the requirement that the causal factor (action) must occur either prior or simultaneously with effect factor(outcome). However, the fact that the action precedes outcome does not establish causal relationship. It might be a mere coincidence that sales training took place prior to increase in sales of the table fans. There is also a possibility that sales training and increase in sales of table fans are strongly associated. This however, does not prove the existence of causal relationship.

The third condition for inferring causation is the absence of other possible causal factors. This means that all other possible factor influencing the outcome (in our case increase in sales of table fan) are either absent or are kept constant. It is only then we could say logically that the sales training has resulted in increase of sales of table fans. In reality, it is impossible to find the absence of other factors or to hold some factors constant. For example, we know that the sales of table fans is influenced by weather conditions. Is it possible to keep weather conditions constant? Or can we be sure that the competitor would not change the price? The obvious answer is "No".

In a case where the outcome is completely determined by only one causal factor, we can say that causal factor is the deterministic cause of the outcome. That is the 'causal factor' in this case is both necessary and sufficient condition for the occurrence of outcome. However, in a situation where the outcome is influenced by a host of causal factors, any of the causal factor is the probablistic cause of the outcome. That is to say it is a necessary but not a sufficient condition for the occurrence of outcome.

There are three possible ways to control the influence of extraneous variables. Firstly, we may physically control the extraneous variable. For example, a company trying to study the impact of two different packagings of a product on sales may control the

extraneous variable like price by keeping it constant for both packaging containing the same amount and quality of good. The second way to control the effect of extraneous variables if the physical control is not possible is to randomize the assignment of treatments to test units. The third way to control the extraneous variables is through the use of experimental designs, the discussion of which would follow in the subsequent sections. If the control of extraneous variable on the dependent variable is not possible by any one of the method, we say that experiment is confounded and such an extraneous variable is called a confounding variable.

Activity 1

Suppose the manufacturer of a particular brand of a desert cooler decreased the price of his sets by 5%. It was observed that there was an increase in sales during the succeeding four months as compared to what the company had prior to price reduction. Has the price reduction increased the sales? Justify your answer.

3.4 EXPLORATORY RESEARCH DESIGN

Ideally all marketing research projects must start with an exploratory research as this helps in providing a sharper focus of the situation and a clearer definition of the problem at hand. The exploratory research design, as the name suggests, involves getting a feel of the situation and emphasises a discovery of ideas and possible insights that may help in identifying areas of further rigorous study. For example a food product manufacturer, wanting to introduce a breakfast cereal may be in knowing the desirable attributes of such a product before really defining the product concept. The main objective of the exploratory research is to fine tune the broad problem into specific problem statement and generate possible hypotheses. It therefore, gives useful direction for further research. The exploratory studies are mainly used for:

1) Providing information to enable a more precise problem definition or hypothesis formulation.
2) Establishing research priorities.
3) Giving the researcher a feel of the problem situation and familiarising him with the problem.
4) Collecting information about possible problems in carrying out research, using specific collection tools and specific techniques for analysis.

Since exploratory studies are not conclusive studies, the design of the study is highly flexible and informal. However, rarely ever does formal design exist in case of exploratory studies. Structured and/or standardised questionnaires are replaced by judgement and intuitive inference drawing on the basis of collected data. Convenience sampling rather than probability sampling characterises exploratory designs. The generally used methods in exploratory research are:

a) Survey of existing literature
b) Survey of experienced individuals
c) Analysis of selected case situations.

Survey of existing literature

Published literature presents a very economical source of study for the purpose of hypothesis generation and problem definition. A large variety of published and unpublished data is available through books and journals, newspapers and periodicals, government publications and individual research projects as well as data collected by trade associations. A lot of data is also generated internally in the company and some of it could be relevant to given problem situations. You will read more about types of secondary data and their use in marketing research in the unit on Data Collection. While survey of existing literature may not provide solutions to the research problem, it can certainly provide direction to the research process.

Survey of experienced individuals

Clue to solving many a marketing problems can be had by talking to individuals who have expertise and ides about the research subject. These individuals could be top executives, sales managers, sales men and channel members who handle the product or related products and consumers or potential consumers. The information collection exercise does not involve a scientifically designed survey, it is merely an attempt together all possible information about the subject of research from people who have specific knowledge about it. The success of this type of experience survey depends upon the freedom of response given to the respondent as well as upon the expertise and communication ability of the people questioned. However, at this stage, since the researcher also has very limited experience with the research problem, he may not be able to elicit very valuable responsers from the individulse.

Analysis of selected cases

This method involves the selection of a few extreme examples reflecting the problem situation and a thorough analysis of the same. In certain cases this sort of study may help in identifying the possible relationships that exist between the variables in a given marketing problem situation. The emphasis in this type of study is upon understanding the research subject as a whole. For example, a company is interested in finding out the reasons for the wide variation in sales productivity of its salesmen, as an exploratory study it could thoroughly analyse the case of some of its best and some of its worst salesmen. This exercise may help in identifying possible relationships between demographic and /or personality variables which may affect variation in sales productivity. The relationships, their extent and direction can then be measured using conclusive research designs.

Activity 2

Study three marketing research projects conducted in your organisation or any other organisation that you are familiar with:

1. What sort of preliminary research preceded these projects:

   A) .....

   B) .....

   C) .....

2. What was the type of exploratory design used for this preliminary research.?

.....

.....

.....

.....

.....

.....

.....

3.5 DESCRIPTIVE RESEARCH DESIGN

Primarily in use for preliminary studies, this type of designs are employed to facilitate description and inference building about population parameters and the relationship among two or more variables. Description or inference could be quantitative or qualitative in nature. Descriptive designs only describe the phenomenon under study attempting to establish a relationship between factors. The data collected may relate to the demographically or the behavioural variables of respondents under study or some situational variables. For example, descriptive research design could be suited to measure the various attributes of successful sales people, or evaluate a training programme or a retailing situation.

The design could be used to study how customers behaved when a new sales promotion programme was introduced. It does not, however, determine the extent of association between the different variables i.e. the income and age of people as associated with response to the sales promotion. Descriptive design can however, be used to draw inferences about the possible relationships between variables.

As descriptive designs are aimed only at providing accurate descriptions of variables relevant to the problem under consideration, they are generally used for preliminary and explorative studies. Very often however, the decision makers choose to accept descriptive data which would permit inference drawing about causality between variables. They may not want to or may not be able to afford experimental studies in terms of time involved and as such descriptive design may at times be used for conclusive studies also.

Descriptive designs are a very frequently used, perhaps the most commonly used category of research design. In short descriptive research can be used for the following purposes.

a) To describe the characteristics of certain groups of interest to the marketer e.g. users of the product, potential users, non users, possible receivers of promotional communication by the company and so on.

b) To estimate the proportion of people in a given population who behave in a certain way for example the proportion of consumers who are prone to deals.

c) To make specific predictions for specified future periods.

d) To develop inferences whether certain variables are associated, for example income and shopping place preference.

In comparison to an exploratory research study, the descriptive study is more formal and less flexible. As the descriptive design in directed at collecting qualitative and quantitative data to enable, description of variables, it involves formulation of more specific hypothesis and testing them through statistical inference of discipline designs may include case research designs, longitudinal and cross sectional designs as well as focus group studies. Of these we shall discuss the panel, cross-sectional and focus group designs.

Panel Research Design

Also known as longitudnal research the panel design involves the continual or periodic information collection from a fixed panel or sample of respondents. The elements of this panels may be retail outlets, dealers, consumers or just individuals. A panel is expected to be constant over time and adjustments are made to provide for dropouts and **representativeness**. A continual measurement of variables relating to the elements of the panel provides a very good idea of their response patterns and may provide clues for future behaviour. The longitudal analysis used involves repeated measurements of the same variables to facilitate a variety of inferences to be drawn about the behaviour of the elements of the panel. Recently the omnibus panel has also been devised where though a fixed sample of elements is maintained but the information collected from the sample members may vary over time or vary between individuals. For example, at one time information on attitudes with respect to a proposed packaging change may be

sought while at another time the information on attitudes with respect to a proposed packaging change may be sought while at another time the information may pertain to respondent reaction to an advertising copy.

Though relatively new in India, panel studies are being increasingly used by research agencies to enable a continual tapping of consumer attitudes, developing buyer profiles, brand behaviour in the market and predicting changes in market position of different brands.

Cross Sectional Design

The cross sectional study is aimed at taking a one time stock of the situation or the phenomenon in which the decision maker is interested. Cross sectional designs give the picture of the situation at a given point of time. For example, if the marketer is interested in knowing the market position of the company vis-a-vis its competitors or the number of people who recall the advertisement of the company out of the total population of viewers, a cross sectional design could be employed to gather information Since the data gathered in this type of design consists of responses from a sample which contains a large number of sources, it provides a 'Cross section" of the situation. In quite a few marketing situations cross sectional data furnish good inputs. Consider the following examples:

- An instant food manufacturing company has provided display racks for its retailers so that its product can be prominently displayed. Before embarking upon a second round of distribution of these racks, it wants to find out how many retailers are actually using the racks and the type of retail establishments that are using them.
- An electronics goods company which provides high quality after sales service to its customers and wants to evaluate the customer response to it.

None of the examples given above seek to establish causal relationships. Cross sectional studies may however uncover relationships, which may be conclusively tested by using the experimental designs.

The cross sectional studies may have an informal design rather than an explicitly described design. When the latter is the case, the design only includes the nature and the source of data, the analytical methods and the nature of the expected results.

Focus Groups

The focus group, an informal design, basically consists of bringing together a group of people to have an extensive, free flowing discussion relating to their experiences or opinion.

The usual procedure is to tape record the conversations and later analyse them. The discussion in the focus group is led and guided by a moderator, who has a fairly standardised way of establishing a relaxed and congenial atmosphere, initiating the discussion and keeping it on the desired focus. The purpose of the focus group design is exploratory, seeking to get respondent opinions of experiences on a particular subject matter. The objectives of the focus group study are clearly specified, but there is no structure of the discussion pattern. In fact, a distinct effort is made to keep the discussion free flowing. The success of the focus group design depends upon a well trained and disciplined moderator who can keep the discussion on to the desired course and elicit responses from all the participants of the group. Consider the following example.

An established toilet soap manufacturer wants to change the advertising theme for his prestigious brand of soap, so as to strengthen its image of a luxury soap using a focus group of young women college students, housewives and some executives. A free wheeling discussion about expectations from toiletries in general and bathsoap in particular was generated and recorded to give valuable ideas and insights for the creative strategy to be used for the advertising theme.

Focus groups serve as excellent tools for preliminary research as they reveal problems and opportunities which emanate from the free and frank discussions by the consumers about their problems, fancies, expectations, satisfactions and dissatisfactions. They are easy to organise and interesting to analyse, and become very good sources of ideas for campaigns, product improvement or repositioning, desirability of deals, dealer selection and so on. The danger associated with focus group lies in the possibility that once analysed, the focus group data may be used for conclusive purposes, to draw conclusions or how people behave. The size of the focus group being very small, it is dangerous to draw conclusions about consumers' behaviour only on the basis of the recorded and duly analysed discussions.

Activity 3

Which of the following are true about a research design:

		T	F
a)	A research design helps to keep the study execution in line with the original purpose and intentions.	☐	☐
b)	It tells the researcher what to do at every stage of the execution.	☐	☐
c)	It forms the basis of understanding between a client and an agency.	☐	☐
d)	It is immaterial whether the design is written or is kept informally in mind.	☐	☐
e)	It must include a quantifiable model of the plan of analysis.	☐	☐

3.6 FACTORS INFLUENCING EXPERIMENTS VALIDITY

Before taking up the discussion on Quasi-Experimental Designs and Experimental Designs, it would be worthwhile to understand the factors that influence experiment's validity. There are two important concepts of validity that are related to experimentation namely internal validity and external validity.

Internal Validity: It means that no other plausible cause of the observed results should exist except those tested. This is the basic minimum that should be present in an experiment before any conclusion about treatments can be made. The presence of other extraneous factors or the absence of internal validity, makes the experiment confounded.

External validity: It relates to the question of "generalization" of the results obtained from the experiments. We are interested in knowing whether it is possible to project the conclusions of the experiment to what population, geographical areas etc.

Although, we would want an experiment to satisfy both internal and external validty, it is not uncommon to find experiment meeting all the internal validity requirements and yet be invalid externally. This could possibly happen because of the following reasons:

i) Environment at the time of test may be different from the environment of the real world where the decisions are taken.

ii) Results obtained in a six-week test may not hold good in an application of, say, twelve months.

iii) The population used for experimentation of the test may not be similar to the population where the results of experiment could be applied.

iv) Treatments in the test may be different from the one used in the real world.

It is desirable to understand the sources of errors that are responsible for reducing experiment's internal and external validity. This will help the researcher guide the choice of design to use.

An experiment's validity is influenced by a host of extraneous factors. These are explained below:

1) **History:** History deals with the occurrence of specific nonrecurring events that are external to the experiment but occur at the some time as the experiment. Suppose we measure the sales of a territory before and after salesmen were trained The difference in the sales may indicate the treatment effect (sales training). However, an increase in sales can not be attributed to sales training alone . An improvement in business condition before and sales training might have resulted in the increase in sales. Therefore, longer the time period involved between observations, greater is the chance that history will confound an experiment of this type and hence will significantly affect the results.

2) **Maturation:** Maturation is similar to history except this it is concerned with the changes that occur with the passage of time in the people involved in the design. In our sales training examples, an increase in sales may be due to the fact the salesmen have become more experienced with the passage of time and hence they know their customers better. It may be noted that it is not only people that change over time, but also stores, geographic regions, and organisations etc. The duration of time period has a direct impact on results of maturation effect.

3) **Testing:** In the process of an experiment, respondents may know that their behvaviour is being observed or the results are being measured. This would sensitize and bias respondents. In our example of training salespersons, if the salespersons know that they are being trained to know how effective the training is, they might respond differently than what they would have done otherwise. For example, if sales in a territory are recorded before training of the salespersons and another observation on sales is taken after the training and if the respondents are aware of it, they are likely to become sensitized and behave differently.

   As another example, consider a case where the respondents have filled a pre-treatment questionnaire. The respondents are likely to respond differently if they are asked to fill the same questionnaire after they are exposed to the treatment. This is because they have now become 'experts' with that questionnaire.

4) **Instrumentation:** This refers to the effect caused by changes in measuring instrument or process during the experiment.

   If difference in total rupees sales volume per territory on a "before" and "after" training is used as the yardsticks to know the effectiveness of sales training, a price change during this time interval could make a substantial difference. This is one example of changing instrument. The experience of the investigators, a change in the mood of the investigators, a change in the investigators are some other examples which may affect the measurment and hence the interpretation of the results.

   Selection Bias: The selection bias occurs when test units are selected in such a way as not to be representative of the population. It refers to assigning of test units to treatment groups in such a way that the groups differ before being subjected to treatments. If groups are selected on the judgement of the researcher, there is always a possibility of selection bias. The selection procedure should be random as this will result in a measureable random variation. However, if a non-random procedure is adopted, the results will be affected by a non-measurable error.

   Test Unit Mortality: It is quite possible that some of test units might drop out from the experiment while it is in progress. Referring to the sales training example, some of the sales persons, who are undergoing training might quit before the training is over. There is no way of knowing whether it is those who were not improving left the training. Also, it is not possible to determine whether those who left the training would produce the same results as those who stayed till the completion of the training programme. In short, the interpretation of the results will become difficult task.

Activity 4

Explain with the help of suitable examples the factors that influence experiment's validity.

.....

.....

.....

.....

.....

.....

.....

3.7 QUASI-EXPERIMENTAL DESIGNS

In order to facilitate the discussion on Quasi-experimental Designs and Experimental Designs (to be discussed in 3.8), the following notational system will be used.

i) X represents the exposure of a test group to an experimental treatment whose effect is to be observed or measured.

ii) O refers to the measurement or observation of the dependent variable on the test units. Suppose more than one observation or measurement is taken, the subscripts O_1, O_2, O_3 etc., will be used to indicate temporal order.

iii) R symbolizes that individuals have been selected at random to separate treatment groups or groups have been allocated at random to separate treatments.

iv) Movement from left to right indicates a time sequence of events.

v) All notations in one row indicate that subjects belong to that specific treatment group.

vi) Vertical arrangements of the notations mean that the events occur simultaneously.

To make the notational scheme clear, let us consider the following examples:

Example 1

The symbols $O_1 \times O_2$ represents that the measurement were taken on a group both prior to (O_1) and after (O_2) of receiving the treatment (X).

Example 2

The symbols

$R\ O_1 \times O_2$

$R \quad \times O_3$

Indicate that two groups of individuals were assigned at random to two different treatment groups at the same time. The groups received the same treatment (X). One group received both a pretest (O_1) and post test measurement (O_2). The second received the post test measurement (O_3) at the same time as O_2.

Now having understood the notation scheme, we begin our discussion on Quasi-experimental designs.

Quasi-experimental Designs: In these designs the researcher has control over data collection procedures but lacks control over the scheduling of the treatments and also lacks the ability to randomize test units' exposure to treatments. There are various designs which fall under the category of quasi-experimental designs. Some of these will be disoussed here These designs have inherent weakness as their internal validty is questionable They lock the control attributes of the truly experimental designs. The

following quasi-experimental designs will be discussed in this section.

1) After-Only without Control Group
2) Before-After without Control Group
3) The Static-Group Comparison
4) Logitudinal Designs (Time Series Designs)
5) Multiple Time-Series Designs.

1) **After-Only without Control Group:** This design is also called a "One-shot case study." The design may be presented symbolically as follows:

X O

In this design, test units are not selected at random. The test units are either self-selected or arbitrarily selected by the experimenter. A single group of test units are first exposed to treatment (X), and then measurements are taken on the dependent variable afterwards (O).

For example, a sales manger may request the members of his sales force to volunteer participation in the sales training programme and their sales performance is measure after training programme is completed.

It is not possible to draw meaningful conclusions from such a design. This is because we do not have any observation (measurement) on the sales performance of the sales persons prior to their undergoing training.

The level of O is the result of various uncontrolled factors. The effect of history, maturation, selection and mortality are substantial and non measurable making the design internally invalid.

2) **Before-After without Control Group:** This design is also termed a "One group pretest-post test design." This design is superior to the design" After-Only without Control Group." It provides a measurement of what existed prior to the test group being subjected to the treatment. This design is presented symbolically as follows.

O_1 X O_2

Referring to the sales training example as discussed in "After-only without Control Group", a pretest measurement of sales performance has been added to the design discussed earlier. Now the question is whether the difference is sales performance as measured by $O_2 - O_1$ can be attributed to the sales training programme? One could list out a number of extraneous factors which could be responsible for this difference ($O_2 - O_1$) thereby making this design useless and disqualified for internal validity. Some of these factors are listed below:

i) The test units had volunteered for the sales training programme (selection bias)
ii) Economic environment might have improved during that period (history)
iii) The salespersons might have gained more experience during the training period which would have resulted in the improved sales performance (maturation).
iv) The pretest measurement might have affected performance(testing)
v) Prices might have changed during that period (instrumentation)
vi) Some test units might have left during the period of training (moratality)

3) **The Static-Group Comparison Design:** In this design we use two groups of subjects. Group 1 is exposed to the treatment whereas Group 2 has not been

exposed to it. Group 2 is called a control group and is used as the baseline for comparison since it has not been exposed to the treatment. In marketing, the control group treatment is called the current level of marketing activity. The design may be diagrammed symbolically as follow:

Group 1 (Experimental Group) : X_1 O_1

Group 2 (Control Group): X_2 O_2

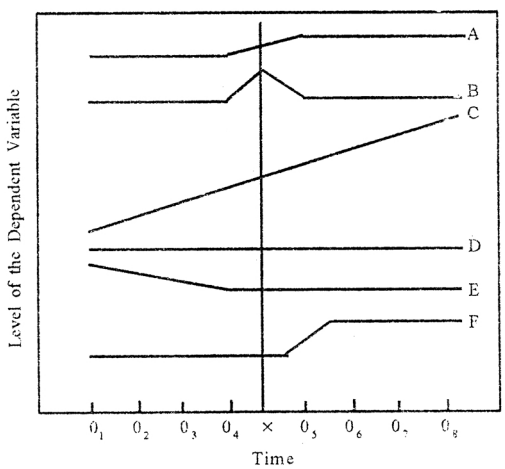
Here X_2 represents baseline marketing programme (existing programme) which is to be compared with the new programme (treatment) X_1 . For example X_1 could be new sales training programme which is to be compared with the existing programme X_2 (may be called control group treatment).

The experiment result O_1-O_2 could be attributed to the selection procedure adopted for two groups. The test units were not randomly selected. So the result may be attributed to the pretest difference caused by selection procedures. Test unit mortality could also be another probable cause. These two reasons put together could make this design invalid.

4) **Longitudinal Designs (Time-Series Design):** This is an extension of one-group pretest-post test design. Here we take periodic measurements for some test units. The treatment is introduced and the measurement on the test units are again taken at various points of time.

For example, a company may be interested in studying the market share of one of its products. It takes measurement on the market share at periodical intervals. Then a new advertising compaign is introduced and periodic measurements on the market shares are taken again.

In order to explain this design, let us consider the following figure and explain the results of this type of design.



Possible Results from a Time-Series Experiment

At the outset, this design may appear to be similar to one group pretest-post test design. However, there is a difference between the two designs. In the pretest-post test design we would have obtained only two measurements namely O_4 and O_5 Here in this design we have in fact taken many pretest-post test observations so that there is more control over extraneous variables. In the pretest-post test design, where we use only two observations O_4 and O_5, the analysis of various situations i as follows:

a) The results for situations A, B and C are identical namely that the market share has increased.

b) For situations D, E and F, the conclusion is that the new advertising compaign (X) has no impact on market share as it has remained constant.

The conclusions arrived at from the various situations of the Longitudinal designs are as follows:

i) In situation A, the compaign has had a short-run positive effect, after which the market share sustained.

ii) In situation B, it is observed that the rise is market share was temporary. The new compaign had a short-run positive effect. We note that the market share reverted to the level existing before the application of the treatment.

iii) In situation C, we observe that the changes that occur after the application of the treatment are a continuation of a trend (they also existed prior to the application of the treatment). Therefore, it is quite likely that the new advertising compaign had no effect on the market share.

iv) In situation D, the conclusion would be same as obtained for situation C.

v) In situation E, it appears that the new advertising campaign seems to have stopped the decline in market share.

vi) In situation F, the treatment (the new advertising campaign) had a delayed effect and therefore it required a longer period to appear.

Because of multiple measurements in this design, the effect of maturation, testing and instrumentation can be ruled out. Further if the test units are selected randomly and measures are taken to prevent panel members (test units) dropping out, one could atleast reduce selection and test unit mortality factors. However, the main weakness of this design is that here it is not possible to control history. However, the researcher could maintain a record of various events like unusual economic activity, economic changes etc. and if no changes are observed, one could reasonably conclude that treatment has had an effect.

5) **Multiple Time-Series Designs:** This designs may be diagrammed as:

Experimental Group	$O_1\ O_2\ O_3\ O_4$	X	$O_5\ O_6\ O_7\ O_8$
Control Group	$O'_1\ O'_2\ O'_3\ O'_4$		$O'_5\ O'_6\ O'_7\ O'_8$

Here the experimental group is the one which was selected for sales training (X) and the observation O_1, O_2, ... represents the sales volume of this group. A comparable group or the sets of the salesmen could be treated as control group. Their measurements on sales volume are symbolised as O'_1, O'_2, O'_3 ..... etc. After training, the measurements on the sales volumes of both the groups are taken. The effect of the sales training may be obtained by comparing the average sales volume for the two groups before and after treatment. The main weakness of this design is the possibility of an interactive effect in an experimental group.

Activity 5

Describe a few specific situations from the functional area of marketing where you would recommend each of the following designs as being best suited . Justify your answer.

a) After-Only Without Control Group

b) The Static-Group comparison

c) Longitudinal Designs

.....

.....

.....

Activity 6

Compare and contrast One Group Pretest-Post Test Design with Longitudinal Design. Use a suitable example from marketing.

.....

.....

.....

3.8 EXPERIMENTAL DESIGN

True experimental design provide a stronger and more reliable basis for the existence of casual relationship between variables. Here, the researcher is able to eliminate all extraneous variables from the experimental and the control group through the use of a random selection procedure. One of the advantages of using random selection procedure is that we can use inferential statistical techniques for the analysis of experimental results. One such technique is the analysis of variance. The following experimental designs will be discussed in this section.

1) After-only with One Control Group
2) Before-After with One Control Group
3) The Solomon Four Group Design
4) Completely Randomised Design
5) Randomizaed Block Design
6) Latin Square Design

1) **After-Only with One Control Group:** Here, two groups are experimental and one control group are set up by randomisation procedure. The design requires only one treatment and post test measurement of both of experimental and the control group. The design may be written symbolically as follows:

Experimental Group	R	X	O_1
Control Group	R		O_2

It may be noted that no pretest measurements are taken in any of the two groups. The difference between the post test measurement of the two groups is taken as a measure of treatment effect. Here O_1 measurement is composed of treatment effect and all the other extraneous variables and O_2 is composed of extraneous factors only. Therefore :

$O_1 - O_2$ = (Treatment effect + Extraneous factors) - (Extraneous factors)
= Treatment effect.

Here the absence of pretest measurement provides control over the testing and instrument effect. Also the interactive testing effect cannot occur. We, however, assume in this design that the pretest measurement of the characteristics under study are same in both the groups. It is also assumed that the test unit mortality affects each group equally. These assumptions can be justified if we take large-randomised samples. This design is very widely used in marketing.

2) **Before-After with One Control Group :** This design may be presented symbolically as:

Experimental Groups	R	O_1	X_1	O_2
Control Groups	R	O_3		O_4

This design is also called Pretest - Post test Design.

In this design two groups — one experiment and one 'control' — are set up by randomisation process. The extraneous variables operate equally on the experimental

and control groups. Pretest measurement are taken in both the groups. However, it is the experimental group which is exposed to treatment. Post test measurement are taken for both the groups.

Since the extraneous factors operate equally on both the groups. therefore, the only difference between the two groups is as a result of the treatment of the experimental group.

Let A = Change in the measurements of the experimental group

$$O_2 - O_1$$

Let B = Change in the measurements of the control group

$$O_4 - O_3$$

Therefore, treatment effect - A-B = $(O_2 - O_1)-(O_4 - O_3)$

The main problem with this design is that it is not possible to remove sensitizing testing effect from the measurement effect. The estimate of the treatment effect given above also includes the sensitizing testing effect. The sensitizing test effects occurs in the experimental group and not in the control group. The individuals in the experiment group get affected because of the pretest measurements and may become very cautious and behave differently after being exposed to the treatment. Therefore the measure $A = O_2 - O_1$ will include a sensitizing testing effect. Since the control group is not exposed to the treatment and therefore this is missing in $B = O_4 - O_3$. Therefore, this shows that the treatment effects as given by A-B will include sensitizing testing effect and therefore we cannot generalise the result of our treatment.

3) **The Solomon Four Groups Design :** This design is also called the four group six study design. This test overcomes the limitations of pretest — post test control group designs. This design is infact a combination of 'Before After with one control group' and 'After only with one control Group' designs. The design is symbolically presented as:

Experimental Groups :	R	O_1	X	O_2
Control Groups 1 :	R	O_3		O_4
Experimental Groups :	R		X	O_5
Control Groups :	R			O_6

You may note that the experimental group 2 and control group 2 do not receive any pretest measurements. The results of group 2 help us in measuring and eliminating the sensitizing testing - effect since this type of sensitizing cannot occur in measurement O_5. The results of the difference of various pretest and post test measurements give us the following effects.

Experimental Groups	:	$O_2 - O_1$ = treatment effect + extraneous + sensitizing testing effect)	(i)
Control Groups 1	:	$O_4 - O_3$ = extraneous factors	(ii)
Experimental Groups 2	:	$O_5 - O_1$ = (treatment effect + extraneous factors)	(iii)
		$O_5 - O_3$ = (treatment effect + extraneous factors)	(iv)
Contol Groups 2	:	$O_6 - O_1$ = extraneous factors	(v)
		$O_6 - O_3$ = extraneous factors	(vi)

taking the average of equations (iii) and (iv), we get

$$O_5 - \frac{O_1 + O_3}{2} = \text{Treatment effect + Extraneous effect} \quad \text{(vii)}$$

By taking the average of equations (v) and (vi), we get

$$O_6 - \frac{O_1 + O_3}{2} = \text{Extraneous effect} \quad \text{(viii)}$$

Treatment effect

$$= \left(O_5 - \frac{O_1 + O_3}{2}\right) - \left(O_6 - \frac{O_1 + O_3}{2}\right)$$

$= O_5 - O_6 =$ Treatment effect

By substracting equation (vii) from equation (i), we obtain the sensitizing testing effect as:

$$\text{Sensitizing testing effect} = (O_2 - O_1) - \left(O_5 - \frac{O_1 + O_3}{2}\right)$$

Please note that in order to measure the change in experiment group 2, we required an estimate of pretest measurement on O_5, we assumed that four groups are equal before the experiment because of the random assignments of the subjects to all four of our groups. Two pretest measurements O_1 and O_3 were taken on experimental group 1 and control group 1 respectively. Given the just mentioned assumption, the pretest measurement of experimental group 1 and control group 1 could be taken as an estimate of pretest measurement of experiment group 2. That is what we used in the equations ((iii), (iv), (v) and (vi) listed above).

This design is also referred to as "the ideal control experiment." As you have seen that we are not only able to control all extraneous variables and sensitizing testing effect, but also could get their estimates. However, this design is not very commonly used in marketing because of increase in cost, time and effort required to conduct the experiment.

The three experimental designs discussed above allow us to manipulate or control an independent variable so as to measure its effect on the dependent variable thereby enabling us to make proper casual statements. Below we will discuss some designs where we can manipulate or control more than one level of independent variable on the dependent variable.

4) **Completely Randomised Design:** The test is useful when the investigator is interested in studying the effect of one independent variable on the dependent variable. The independent variable should have a number of categories (nominal scale). Each category of the independent variable is considered as a treatment. Suppose we wish to sell a product and want to know what price should be charged. Here the independent variable is price and the three different levels of price namely low, medium and high are used as treatments. The test units (shops) are randomly assigned to three different treatments (price levels). Suppose our dependent variable is sales, we could always compare average sales level corresponding to each of the treatment to determine which treatment (price level) was the best.

The limitation of this design is that it does not take into account the effect of various extraneous variables on the dependent variable. In the example which we have considered, the possible extraneous factors would be competitor's price, size of the shops, the price of the substitute of the product in question and so on. It is assumed in this design that the extraneous factors have the same impact on all the test units.

5) **Randomized Block Design:** The randomized block design is an extension of completely randomized design. In the completely randomized design it was assumed that all the extraneous factors have the same impact on all the test units. This may not be true.

In the examples used in the completely randomized design, we had three price levels tested in various shops. We did not take into consideration the fact that the

differences in the result could be due to the differences in the size of the shops. One could classify various shops on the basis of their size as:

Small size shops
Medium size shops
Large size shops

In the randomized block design it is possible to separate out the effect of one extraneous factor from the results thereby providing a clearer picture of the impact of treatment on sales.

In this experiment, we may for example, randomly assign nine small size shops to the three price level in a way that there are three shops for each price level. Similarly one could assign randomly nine medium size shops to the three price levels and nine large size shops to the three price levels. Having done this, one could use analysis of variance technique to separate out the effect of extraneous variable (size of shops) from the total experimental error and also analyse the effect of treatment (price level) on sales (dependent variable)

6) **Latin Square Design:** This design allows the researches to control and measure the effect of two extraneous variables on the dependent variable. Here the two extraneous variables need to be divided into as many categories as the number of levels of the independent variable (treatment). The Table 1 below illustrates the layout of Latin Square Design.

Table 1: Latin Square for Various Advertising Campaigns

		Packaging			
		I	II	III	IV
	1	A	B	C	D
Cities	2	B	C	D	A
	3	C	D	A	B
	4	D	A	B	C

Suppose a marketing organisation wants to study the impact of four types of advertising campaigns on the sales of a product. The sales could be affected by say two extraneous variables like the type of city and the type of packaging. If Latin Square Design is to be used the cities need be divided into four types namely 1, 2, 3, 4 and four different types of packaging say I, II, III and IV are required. In Table 1 we have identified four categories for the row variable, the type of city, and four categories for the column variable, the type of packaging. The treatment variable the types of advertising campaign is identified as A, B, C and D. We note that the number of categories of extraneous variables is the same as the number of treatments (4).

The treatments (four advertising compaigns) are assigned randomly to the cells of the table in such a way that each treatment appears once and only once in each row and each column. One this arrangement is made, the analysis of variance technique could be used to separate out and measure the effect of the two extraneous variables in question and also analyse the effect of treatments (four types of advertising campaign) on the dependent variable namely sales of the product.

This design is very complex to set up. Further, it can be very expensive to execute.

Activity 7

How would you use Betore aficr with one couful Grraip on reasore not Wicctiveness of a new advertising campaign on the suics of a pachular buand of skaboci Also acntion the worknces of this design

Activity 8

list out fow maekring studies where Latuu squerc design couud be usrd

3.9 SUMMARY

This unit on research design gives an overiew of the research designs and their application in marketing research. After discussing the meaning and benefits of a research design, the discussion focusses on the four basic types of designs namely exploratory, descriptive, quasi-experimental and experimental. Causality, the relationship between data and conclusions, forms the basis of this classification. The nature of causality and the need to measure it has also been discussed.

The major design categories and their applications have been described. Specific subtypes within the major categories have also been discussed. Under the descriptive design, panel design, cross sectional design and focus groups were discussed. The quasi-experimental designs discussed include after only without control group, before-after without control group, the static group comparison, longitudinal design and multiple time series designs.

Under the experimental design categories the designs covered are after-only with one control group; before-after with one control, Solomen four group design. completely randomized design, randomised block design and Latin square design. Questions of internal and external validity of the experiment have also been discussed.

3.10 SELF-ASSESSMENT QUESTIONS

1) Explain the solomon-four group design. How far does this design succeed in controlling different extraneous variables ? Illustrate your answer with the help of a suitable example from marketing.

2) Explain the various extraneous variables which if not controlled in an experiment may contaminate the effect of the independent variable.

3) What design would you use to

   a) Test whether the coloured newspaper advertisement is more effective than the existing black and white advertisement ?

   b) Know which of the several promotional techniques is most efficient in selling a particular product ?

   c) Determine whether additional shelf space allocated to a product would increase the sales of that product in the supermarket ?

   d) Determine the type of customers who would be the first adopters of a newly introduced brand ?

Justify your answer.

4) What are the marketing research situations suitable for

   a) Focus group study

   b) Panel research design

   c) Cross-sectional design

   d) Quasi-experimental design

5) Distinguish between exploratory and descriptive research design.

3.11 FURTHER READINGS

Freund, John E. and Frank J. Williams, "Elementary Business Statistics — The Modern Approach", Prentice Hall International Edition.

Kinnear, Thomas C. and James R. Tayler, "Marketing Research - An Applied Approach" McGraw-Hill International Edition .

Luck, David J. and Ronald s. Rubin, "Marketing Research" Prentice-Hall of India Pvt. Ltd.

Green, Paul E. and Donald S. Tull, "Research for Marketing Decisions" Prentice Hall of India Pvt. Ltd.

Westfall, Boyd and Stasch, "Marketing Research — Text and Cases" Rechard D. Irwin, Inc.

NOTES

NOTES

NOTES

Uttar Pradesh
Rajarshi Tandon Open University

MBA-3.44
Marketing Research

Block

2

DATA COLLECTION

UNIT 4

Data Collection 5

UNIT 5

Sampling 21

UNIT 6

Questionnaire Design and Development 41

UNIT 7

Attitude Measurement & Scaling 53

Expert-Cum-Course Writers Committee

Prof. Rakesh Khurana, *Chairman*
School of Management Studies
IGNOU, New Delhi

Prof. Abraham Koshy
Indian Institute of Management
Ahmedabad

Prof. Y. L. Grover
Management Consultant
New Delhi

Prof. Deepak Chawla
International Management Institute
New Delhi

Prof. Mukul Gupta
International Management Institute
New Delhi

Mrs. Neerja Wable
Indian Market Research Bureau
New Delhi

Shri Saurabh Khoshla
MD, Tulika Advertising Agency
New Delhi

Prof Sudas Roy
Indian Institute of Mangement
Calcutta

Dr. N. Kinra
Indian Institute of Management
Lucknow

Prof. A.P. Arora
Management Development Institute
Gurgaon

Dr. Sanjay K. Jain
Delhi School of Economics
University of Delhi, Delhi

Mr. B. Narayanaswamy
Executive Director
Indica Research, New Delhi

Dr. P. S. Vishwanathan
Hindustan Computer Ltd.,
New Delhi

Prof. Sharad Sarin
XLRI
Jamshedpur

Prof. Mohan Aggarwal
XLRI
Jamshedpur

Mr. D. K. Bose
Media Director
R. K. Swamy, BBDO, New Delhi

Dr. (Mrs.) Madhulika Kaushik
Director
School of Management Studies
IGNOU, New Delhi

Prof. R. K. Grover
School of Management Studies
IGNOU, New Delhi

Mr. Qasemuddin Haider
Asst. Regional Director
IGNOU, Patna

Mr. Ravi Shanker
School of Management Studies
IGNOU, New Delhi

Dr. Kamal Yadava
Course Coordinator
School of Management Studies
IGNOU, New Delhi

Print Production : B Natarajan, Copy Editor, SOMS, IGNOU

September, 2000 (Reprint)

© *Indira Gandhi National Open University, 1998*
ISBN-81-7605-337-6

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the Indira Gandhi National Open University.

Further information on the Indira Gandhi National Open University courses may be obtained from the University's Office at Maidan Garhi, New Delhi - 110 068.

Reproduced and reprinted with the permission of Indira Gandhi National Open University by G. S. Shukla, Registrar, U.P.R.T.Open University, Allahabad (December 2011)
Printed by . Chandrakala universal Put. Ltd. 42/7 Jawahar Lal Nehru Road Allhabad 211002

BLOCK 2 DATA COLLECTION

Block 1 of this course gave you an overview of Marketing Research, its use and relevance in Indian Context and research design. This block takes you a step further and discusses various issues & concepts related to data collection. The block consists of four units — Unit 4, Unit 5, Unit 6 and Unit 7.

Unit 4 gives you the details of different types of data, basic methods of data collection and sources of errors associated with primary and secondary data. Unit 5 on 'Sampling' describes the importance of sampling in marketing research studies and explains the process of sampling and issues related to sampling design. Unit 6 titled 'Questionnaire Design & Development' explains the different types of questionnaires and the issues involved in designing a questionnaire. The last unit of this block on 'Attitude Measurement & Scaling' deals with various types of measurement scales.

UNIT 4 DATA COLLECTION

Objectives

After studying this unit you should be able to:

- Define the different types of data needed for marketing research
- Evaluate the merits and demerits of each type of data
- Discuss the various sources of secondary and primary data
- Explain the basic methods of data collection
- Describe the sources of errors associated with secondary and primary data.

Structure

4.1 Introduction
4.2 Data and the Research Process
4.3 Secondary Data — Need and Usage
4.4 Sources of Secondary Data
4.5 Advantages and Limitations of Secondary Data
4.6 Sources of Primary Data
4.7 Basic Methods of Data Collection
4.8 Sources of Error in Primary Data Collection
4.9 Summary
4.10 Self-Assessment Questions

4.1 INTRODUCTION

In the last unit you have studied the research design alternatives available to the researcher. Whatever may be the type of research design chosen, it is necessary to collect accurate and reliable data in order to achieve the research objectives. In this unit, we shall discuss the different types of data used for marketing research problems and their role in the research process. We shall also learn about the major sources of secondary and primary data as well as the basic methods of collecting data. Whether data is collected by communication methods or through observation, errors of different types tend to be associated with different type and tool of data collection. We shall also discuss the important sources of error associated with data collection.

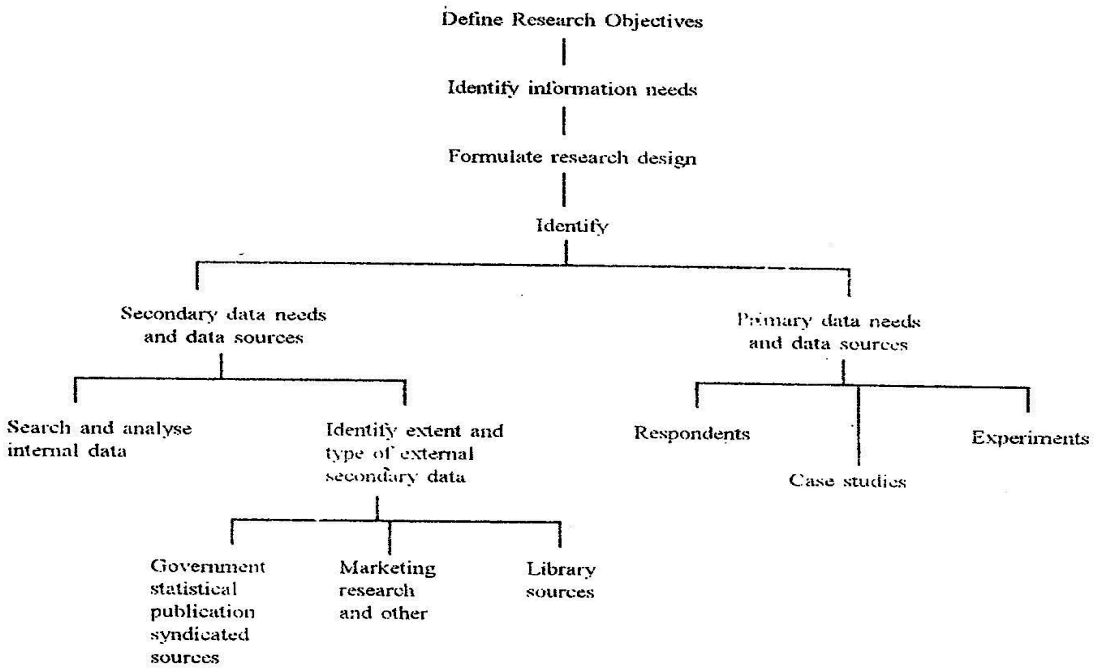
4.2 DATA AND THE RESEARCH PROCESS

Once the research objectives have been defined the next logical step is to identify the nature and type of information needed to achieve these objectives. Fig. 1 below represents the initial steps in the research process with emphasis upon the types of data sources and options available to the researcher. The research objectives define the type and extent of information needed to achieve the research objectives — the data needs are further clarified by the type of research design chosen as well as by the nature of research e.g. whether the problem at hand is that of exploratory or preliminary research or causal and conclusive research and so on.

All data sources available to the researcher can be classified as either secondary or primary. Secondary data are already published data collected for purposes other than the research problem at hand. Primary data are those generated specifically for the purposes of the research problem in question.

Generally, the first issue in the data collection process is to determine whether the information needed for the research problem has already been generated/collected. Only after the assessment of the data already available, additional data needs should be specified.

Figure 1 : Data and the Research Process



Types of Data

As you have already noted there are two general types of data — Primary and Secondary classified on the basis of purpose of collection or source. Primary data are those that are collected specifically for the research situation at hand. Conversely secondary data are already published data collected for purposes other than the specific research needs at hand. On the basis of location of sources, secondary data may again be classified as internal or external data. The data originating within or available within the organisation as a byproduct of the MIS or the routine reporting system is called internal data of any given marketing research problem, initial data collected for purposes other than that specific problem could be termed internal secondary data.

Secondary data generated outside the organisation is termed external secondary data and can be collected from a multitude of sources like government publication, trade association publications, official reports, journals and periodicals and publication of marketing research agencies. Secondary data can also be thought from research agencies though this is a fairly expensive preposition.

Secondary data may also be classified on the basis of whether it is periodic data or adhoc data. Periodic data characterises most statistics collected over fixed periods of time like the census data, data from statistical abstracts of trade and other sectors, price indices and so on. Adhoc data, on the other hand refers to the data obtained from a certain project report. Such data necessitates an external search to find the source from which the data can be availed of.

As the figure 1 shows the major sources of primary data are respondents, analogous case studies and experiments. The secondary data on the other hand can be obtained from official government publication, marketing research agencies and syndicated sources (which sell the data) and the library sources. Published individual research projects have not been dealt with separately as they are included the library source. Let us now examine the data options and their sources in detail.

Activity 1

Talk to the person responsible for marketing research function in your own organisation any other organisation that you are familiar with. Try to find out, with respect to (a) a product introduction decision, and (b) a new sales incentive scheme introduction decision, what were the type of data utilized by the organisation. for these decision What were the major sources of data and the methods used to collect the data

a) Product introduction decision

i) Type of data

.....

.....

.....

ii) Major data sources

.....

.....

.....

iii) Methods used

.....

.....

.....

b) Sales incentive scheme introduction

.....

.....

.....

4.3 SECONDARY DATA— NEED AND USAGE

Before discussing the sources of secondary data, let us first try to envisage what type of research objectives and information needs may necessitate use of secondary data. Some examples could be

a) You want to estimate the total market potential of thermocol packaging cases in the country

b) Being a manufacturer of television sets, you want to develop a national method for establishing sales quota statewise.

c) As a person making automobile batteries, you want to estimate the potential market for battery replacements as well as develop methods for determination of countrywise, statewise and districtwise sales quotas.

d) You would like to predict the potential sales of paper and allied products to wholesaler and retailer, for the year 1991.

e) You are required to select a city in Western Maharashtra to locate another distribution centre for the spare parts of your two wheeler products.

f) You are interested in allocating your sales promotion budget in proportion to the potential markets, statewise in the northern pan of India.

The type of secondary data required to meet some of the above research objectives would be population data, estimates of total television sales, statewise, number of households with televisions per capita income estimates, distribution of income statistics. Number of registrations, number of households without cars, value of box shipments by end use and so on. You have read about the various kind of research designs in the preceding unit. With the type of research design chosen need and use of secondary data differ.

In monitoring research, the researcher is trying to keep a track of a programme or phenomenon with respect to the variables under study (it may be the progress of pan

under study or a specific progress) and is keen to sense changes in the environment which could help or hamper the product programme under study. Most of the data for this type of monitoring would come from secondary sources and quite a substantial part of it would be internally generated. For example accounting data on sales, expenses and margins earned could provide significant clues to the progress of the product in the given market. Sales force reports could generate observations about customers and market trends that could be used to sound early warning bells to the executives.

Preliminary research also makes extensive use of secondary data, though data can be gathered afresh also, in small or informal primary studies. In this case secondary data is used to estimate parameters of a problem or a trend and the changes in its environment. Preliminary research, since it explores nature and extent of the problem, does not generally merit the expense & delay of primary data collection. In this preliminary phase, usually descriptive data, both from internal and external sources can be used to understand and to define the nature of a research problem.

If you are undertaking exploratory research (for example if you are trying to explore the various possible combination of the communication mix which could yield the optimum mix) you will be able to make a relatively li: ed use of secondary data because this type of research is concerned with finding specific alternatives that should considered while arriving at a decision. Of course past experience of the company, its records or internal data could provide guidelines for action as could external secondary data, especially published material on consumer perception and reaction to different tools of the communication mix, media habits and so on.

Conclusive research, concerned as it is with establishing dependence or independence of the variables under study is mainly based on primary data gathered to suit the objectives of the research problem under study. For example, if you are interested in finding out whether a given sales promotion programme has a significant impact on the sales of your new range of moulded luggage then you would have to rely primarily on fresh data generated specifically to establish the cause and effect relationship between the sales promotion programme and the luggage sold.

However, secondary data may be used for comparison purposes even in conclusive research. You might like to compare the industry trend in respect of promotional expenditures and sales with your own company position, or compare the position of the company statewise with that of its select competitors. Also secondary data is useful in framing a primary data collection programme, as well as in framing the sample or determining the sizes of the sub-samples.

4.4 SOURCES OF SECONDARY DATA

Government Agencies and Official Publications: The largest single source of secondary data in macro terms are the publications by the Union government. Marketing researches have relied on this source of data for estimating market potential and sales forecasts, determining distribution penetration and location of intermediate and final outlets, as well as for defining sales territories and routing schedules. Estimates of income and expenditure patterns become good starting points for estimation of paying capacity for different products and services. Estimates of literacy levels become effective inputs in planning promotional strategies. Over a period of time, the variety and depth of government data has increased manifold and it relevance to marketing research function has consequently enhanced. The following sources have special relevance for marketing research.

1) Population Statistics

Population statistics in India have been collected every ten years since 1871-72. Besides forming a basis for most macro planning estimates, census data in respect of the marketing research function is the most important source of national population statistics and its basic characteristics especially demographic and economic. It provides factual bases for estimation of consumer demand for various goods and services, by furnishing

data on size of population and its distribution by age, sex, occupation and income levels. Census also furnishes a sampling frame for reliable sample designs. Estimates of population density, distribution, literacy levels, consumption patterns all become important inputs in distribution decisions, communication decisions and overall marketing policy.

2) Statistical Abstract of India

Brought out annually by the Central Statistical Organisation, this publication contains the statistics of various section of the Indian economy for the preceding five years. Statewise, statistics for these economic variables are also furnished periodically.

3) Estimation of National Product, Savings and Capital formation: (White paper on National Income)

This is also an annual publication of the Central Statistical Organisation. It publishes annual estimates of national income, savings and consumption, capital formation and expenditure as well as national and public sector accounts.

4) Monthly Statistics on the Production of Selected Industries

To bridge the gap between the census taken and data published, the C.S.O. publishes monthly statistics relating to production, installed capacity and stock positions in selected industries. More than 90 industries are included in the compilation of chosen statistics.

5) Basic Statistics Relating to Indian Economy

Published annually by the statistics and survey division of Planning Commission this publication contains basic indicators on various aspects of economy thee past few years based on time series.

6) India, Pocketbook of Economic Information

It is an annual publication of the Ministry of Finance. Giving particular emphasis on estimates of foreign financial and international comparisons, it includes statistics on the various aspects of the national economy.

Other important publications include India, a reference manual. Agricultural Situation in India, RBI Bulletin, Economic Survey. Bulletin of food Statistics, Commercial Crop Statistics and Indian labour statistics (which includes transport and communication statistics, data on employment in shop and commercial establishments etc.).

7) Trade Statistics

The Department of Commercial Intelligence and Statistics compiles monthly statistics of commercial intelligence and statistics, publishes data on import and export of goods in terms of their quantity and value, classified as received from or sent to centres of consignment. In addition, this publication provides information on value of foreign trade, balance of trade, foreign trade with each country and currency area, foreign trade in groups of commodities with each country and currency area, foreign trade with selected countries etc. It therefore, furnishes a good source of data for assessment of international market trends and potential. The Reserve Bank of India also publishes statistics on imports and exports based on exchange control data.

In addition, official reports both published and unpublished from organisations like National Sample Survey Organisation, Directorate of Economics and Statistics, Labour Bureau etc. furnish data on the different aspects of our economy. Agencies like Cottage Industry Board, National Small Scale Industries Corporation, Tea Board, Handicraft Board etc. provide specific information in respect of their own product categories.

Other Sources of Secondary Data

Library: The library sources of marketing data include the whole gamut of publicly circulated material i.e. government documents and reports, books, periodicals, journals, individuals research project reports and trade association publications. As the categories

noted above are intended for public circulation, the library represents an easy, economic and efficient source of secondary data.

Some of the item inventioned above are the results of extensive original research, some a summary of the research of others and some presenting statements of opinions. In comparison to the official statistics described before, these items tend to include more discussion and less statistics which in turn may help you decide upon the suitability of a particular type of secondary data for your own purposes.

Research Agencies and Data Services: The growing demand for marketing data has brought forth several organisation which collect and sell standardized data. Also called syndicated sources, these agencies include the marketing research agencies as well as the data services which in addition to providing standardized data also undertake specific data collection/research projects. In India, advertising agencies have also emerged as good sources of data on readership media habits, attitudinal research and other communication related areas.

The data that can be obtained from these syndicated agencies includes consumer data, retail data, wholesale data, industrial data, advertising evaluation data and, media and audience data.

Leading marketing research agencies like MARG and ORG regularly survey consumers attitude and opinions regarding consumption behaviour and a variety of contemporary issues relevant to marketing.

Activity 2

A) Study the demand estimation process for an existing product in your organisation or any other organisation that you are familiar with

   a) What were the sources of secondary data utilized?

   b) What specific use of secondary data was made in the study?

B) Repeat the exercise for the demand estimation for a new product that is yet to be launched.

.....

.....

.....

.....

.....

4.5 ADVANTAGES AND LIMITATIONS OF SECONDARY DATA

Advantages of secondary data: The major advantage of secondary data is the economy of resources that it offers both in terms of money and time. Primary research involves selecting the sample frame, determining the sample size, choosing the tools of data collection, getting data collection instrument printed in case of field studies as well as editing tabulating and analysing the results, which turns out to be expensive and time consuming. Secondary data on the other hand can be collected by researcher from published or compiled research at very little cost and usually very speedily.

Another important advantage which characterises some secondary data services is that they provide access to information which would not ordinarily be obtainable by an individual organisation. The census of wholesale and retail establishments for example can require them to furnish details of sales, expenses and profit information which would be inaccessible to an individual researcher. Also, as information like this is collected in the usual course of events, the data is less prone to be biased which may be the case when the data is collected with a specific purpose in mind.

Limitations of secondary data: While economy and access to relatively unbiased information are important advantages, effective use of secondary data requires the overcoming of two major difficulties. These are the difficulties associated with fitness of data and accuracy of the data.

The data fitness problem: The problem here refers to the suitability or fitness of secondary data to the research problem at hand. As secondary data have been complied for other purposes, rarely are they completely pertinent to the information needs of the problem at hand. Even when they are relevant, the degree of fitness to a given research problem may be hampered by (a) Units of measurement (b) Class definition and (c) Recency.

Among the most common limitations of secondary data is the variation found in the units of measurement used for a given variable. For example if you want to have a reliable estimate of the disposable income in order to base your market projections on it, you may discover that depending upon the source, income flay have been measured by individual, household or tax returns. In case you are required to use more than one of these sources the difference in the units of measurement would allow the information to be used only as rough approximation. Another common feature of secondary data is the variation in definition of classes that would be found in different data sources or even in different project reports of the same data source. For example definition of a class of literate population may have been defined as those who can read or write in any language, those who have had at least basic schooling, or those who have attained a certain educational certification, in different reports published by the same source or different sources.

Recency of the secondary data also becomes an important determinant of the use of secondary data. Marketing decisions can hardly be based on information that is dated and therefore no longer relevant. The use of census data which is an invaluable source of voluminous information is severely limited by the fact that not only do we have a decennial census, the reports of the census become available only a year or two after the census. Census information collected in 1981 would lose its relevance to most marketing decisions beyond 1985 or so, even though the data contains some of the most vital marketing data on income. consumption, expenditure savings and investment.

The accuracy problem: Once an appropriate source of secondary data has been identified, it is important to evaluate the accuracy of the data in relation to the research problem at hand. The following steps would help in evaluating the accuracy of the secondary data.

Identify the data source: Secondary data may be obtained from original source or from a secondary source which in turn has obtained them from somewhere else. As far as possible data must be collected from the original sources as you may then be able to evaluate the methodology of research that the original researcher has used. When secondary sources of secondary data are used, not only are details of methodology of original source non-available, errors originating at the secondary source may affect the accuracy of the original data. Moreover, secondary sources usually fail to keep up with the updated versions and adjustments reported by the original sources from time to time as more information becomes available to them. Secondary sources of data must therefore be used with caution and after careful scrutiny.

Examine the purpose for which data was published: Sources which publish to promote the interests of a particular group or for political commercial or social reasons must be treated with caution. Similarly data published to promote sales or to carry on a particular propaganda or to promote views of a particular interest group has limited use unless it is carefully interpreted in the light of purpose of publication. Data published anonymously or by an organisation which is on the defensive or under conditions which suggest a controversy or in a form which reveals a strained attempt at frankness or to controvert inferences from other data are generally suspect."

(I.E. Newmers and John H. Nyers 'Business Research: Text and Cases', New York, McGraw Hill, page-42.)

Assess the source in terms of the quality of data expected: In addition to identifying the organisation which collected the data, the ability of the organisation to procure the desired information must be assessed. For example, the department of internal revenue is much more likely to generate reliable information about income and expenditures simply because it has the authority to glean that kind of information.

Evaluation of the methodology of research used by the source in terms of research design, sampling plan data collection procedures and analysis procedures also furnishes important clues to the accuracy of secondary data. If the primary source is not willing to divulge the methodology of research one should be hesitant about using the data as unwillingness to share the details of methodology usually suggests inadequacies in the procedures used.,

4.6 SOURCES OF PRIMARY DATA

As noted earlier, though nearly all marketing research problems make some use of secondary data, both exploratory and conclusive research situations necessitate using a high proportion of primary data. The major sources of primary data include respondents, analogous case situations and research experiments.

Respondents: Respondents represent by far the most important source of primary marketing data. Marketing decisions are characterised by the fact that they always involve in one way or the other, prediction of the behaviour of the market participants— be it consumer, industrial users, marketing intermediaries or competitors. Decisions as diverse as product introduction, price or channel modification, determination of the advertising budget or reallocation of sales territories would require forecasting of the behaviour of one or more of the above groups. The study of respondents therefore characterises most marketing research situations.

The type of information that may be collected from respondents may include data on past behaviour, intentions of likely behaviour, extent of knowledge, attitudes and opinion and socio-economic characteristics and lifestyle data.

Past behaviour is frequently used as a predictor of future behaviour, acting on the premise that there is a relatively stable relationship between past and future behaviour patterns. Information or past behaviour of respondents is sought on past consumption habits, exposure and awareness of communication tools, experience of competitor's vis-a-vis the company's products or services etc., in order to get an idea of the consumption of the product or service.

Intention of likely action are self prediction of planned action to be taken sometime in future. They represent a very frequently sought information from the respondent. However, as intention only represent concurrently valid statements, actual behaviour may get modified by changes over which the respondent has little or no control, for example— price changes or entry of a new attractive competitive brand. The utility of intentions data is far higher as a predictive tool, when intentions, given the conditions of purchase can be assigned probabilities.

Buyer decisions to purchase a product or services depend to a certain extent on what they know about the product or service. **Extent of knowledge** or awareness is therefore another type of information that is sought from respondents. As a marketer of colour television sets, you would like to know what do the buyers know about your product's attributes like picture quality and sound fidelity, extra features available, price etc. vis-a-vis competitor's products, if you think that these are the criteria which have important bearing or purchase decisions.

Attitudes and opinions represents predisposed mental state to act in a certain way. For example, the statement by a respondent that he prefers fresh ground coffee to the instant blends, gives a clue to his likely behaviour when faced with a decision to buy coffee. Attitude research in marketing has found the greatest use in product designs and

advertising. Both qualitative and quantitative techniques are used to collect information on attitudes and opinions. They have been discussed under the section 'Basic Method of data collection'.

Socio-economic characteristics

Whenever in a consumer research situation there is a basis for the belief that some socio-economic characteristics like income, education occupational status etc, are associated with the purchase of a product or service. information is sought on these characteristics. Study on these types of variables may help defining the target market more precisely and arriving at more accurate pricing and promotion decisions.

Life styles are increasingly being used to segment markets reposition products and targeting positioning strategies more closely to the consumer profile. Life style profiles are developed though systematic study of values attitudes, opinions and interests, as well as by income, all of which can be elicited from the respondents through communication methods or a combination of communication and observation methods. You are already familiar with the information needed and methodology used to develop life style profiles through your exposure to Unit 2 of this course.

Primary data from Analogous situations — Case Study: Evolved from the behavioural sciences, case study or case history is in extensive use in marketing research today. Using situations analogous to or relevant to the problem situation, an in depth investigation is carried out to thoroughly study the case situation. The emphasis of the study is on identifying key variables, defining the nature of their relationships and possibly defining the problem/opportunity in order to suggest alternative courses of action for the decision situation. The source is typically relevant in situations where multiple variables interact to produce the problem or the opportunity. Examples would be research problems involving, a) study of changes in sales performance with the entry of a new competitor, b) contrasting performance levels in different markets, c) transitional stages like study of sales territories where the company is altering its distribution channels from indirect to direct sales.

Simulation: A marketing simulation is an incomplete representation of the marketing system or some aspect of the system. It is a relatively new data source and is largely computer based. It involves the study of dynamics of the marketing system by manipulating independent variables like marketing mix or situational factors and observing their influence on the dependent variables. It would therefore need characteristics of the phenomenon and the relationships between variables to be represented as data inputs.

The researcher creating a marketing simulation is required to conceptualized and record the structural component and assign probabilities to represent the behaviour of the components. Depending upon the phenomenon being studied these components could be buyers, retailers, households etc. while the variables which affect how these components would behave could be price charges, advertising expenditure levels, sales promotion programmes, quality of the product and so on.

The constants in respect to phenomenon under study are represented as parameters of the study which can be manipulated by the researcher in order to experiment and explore alternative marketing strategies. The behaviour of the components results in numerical output. An example could be, suppose you want to study the consumer response to varying advertising expenditures, in terms of sales. The components here would be consumers, the variable is a expenditure and sales is the dependent variable. The researcher would assign probabilities of consumers reacting to a particular level advertising expenditure in terms of buying the product and by varying the different expenditure levels, would get the output in terms of sales expected at these levels. The constants could be the size of the advertising budget.

Simulation is a complicated process and its success depends how well has the actual situation (reality) been represented in the simulation (abstraction). It requires the skill of a specialist and is therefore limited in application as far as the routine marketing research situations are concerned.

Experimentation

Experimentation also represents a fairly rich source of primary data and is used mostly to study cause and effect relationships among marketing variables. Marketing experiments are conducted almost like other scientific experiments. One or more independent variables are consciously manipulated or controlled and their impact on indendent variables is studied. All the same effort is made to control those variables which may hamper the ability to interpret valid causality. Various experimental designs have already been discussed in the preceding unit under the subtitle of experimental research design.

Advantages of Primary data

Primary sources usually provide more detailed information than the secondary sources. This is partly because methods of data collection and the tools used can be tailored more precisely to the informational needs of the researcher. This also contributes to the flexibility of analysis for the research purpose at hand.

Terms and units can be more precisely defined and the researcher can choose the appropriate unit for this purpose with a much greater degree of fit since he has greater degree of flexibility in choosing the appropriate unit. In case of secondary data, as the data have been collected for some other purpose, the researcher may not have much of control over the choice of units used in the secondary data source.

The user in case of primary data can judge the degree of confidence that he may place on the data because he has an accurate idea of the tools and methodology used and their limitations.

In addition to the above general advantages, there are other merits which characterise the different methods of primary data collection. We shall study them under the section concerned with tools of data collection. The major limitations of primary data relates itself to costs in terms of time and money involved, availability of skilled investigators or interviewers and the errors that may creep in various stages i.e. sample selection, sample size selection, tools chosen etc. As the errors related to primary data collection require greater attention they have been explained separately in Section 4.8.

4.7 BASIC METHODS OF DATA COLLECTION

When you, in your everyday life seek information on something from people, the only ways of getting it are either asking for it or observing the phenomenon on which you want information. Formal research has merely formalised and systematized these two alternative methods of data collection i.e. communication and observation. Within each category you will find several tools that can be used to affect communication and/or observation. The fact that distinguishes communication methods is that there is a dialogue with the respondent either spoken or written, while the observation methods involve keeping a track of or monitoring the behaviour of the subject matter of the research, without interfering with the same.

Communication methods

Interviews: Interviews in marketing research are by far the most common method of data collection. Interviews may be:

a)	Structured and Direct	Involving the use of a structured formal questionnaire as well as an interviewer (e.g. surveys using questionnaires).
b)	Unstructured and Direct	Not involving the use of a predecided questionnaire, only the interviewer (e.g personal interviews)
c)	Structured and indirect	Where the respondent is asked to respond to a

non-personal ambigous situation which is later interpreted (e.g. projective techniques like word association)

d) Unstructured and indirect — Where the respondent is asked to respond to non-personal ambigous situation but where the interviewer has considerable degree of freedom in modifying/altering the situation.

a) Structured and Direct (Surveys using Questionnaires)

Among the communication methods in use today, surveys involving structured questionnaires are the most extensively used. If you want to know the consumption patterns of a particular product category, reasons for brand choice, the relative influence of different media on a person's decisions, the natural course would be to ask the people themselves. In order to get comparable information from repondents, formal nondisguised questions are predesigned in the form of a questionnaire. Administrating the questionnaire may involve participation of an interviewer, who is instructed to ask the questions in the order given on the form and to ask only those questions. The questionnaire method, with or without the participation of an interviewer is such a prevalent and versatile method of data collection that it was felt that various stages of questionnaire planning and execution warranted a detailed explanation. Unit 15 of this block is therefore devoted entirely to questionnaire planning and execution. You will get a detailed exposure to this tool of data collection in that unit.

The basic advantages of using questionnaires are versatility and economy. The questionnaire method is versatile enough to be tailored to the needs of most marketing research situations. In addition veriables like knowledge, opinions, intentions, motivations and personal habits which do not lend themselves to observation, can only be elicited through questioning. It is also the only method to get information on past events for which records have not been maintained.

Relatively speaking, questioning turns out to be speedier and less costly then observation. Interviewers using questionnaires have a higher degree of control over information gathering activities than do observers, who have to wait for the respondent to perform the action under study. Moreover the lag time between one interview and another is controllable by the interviewer while the wait time between two successive observations cannot be controlled by the observer.

The questionnaire method is however subject to certain limitations also. Important among them are inability of the respondent to furnish information. Even though they may be willing to share information about themselves, many respondents are actually unable to give accurate, information to the questions asked. For example, it might be difficult for you, at a point of time to give the exact reason of choosing a brand of soap over another unless you have carefully analysed these reasons beforehand. Therefore, on questions of buyer motivations quite often inaccurate information is given which interferes with the reliability of data collected.

Unwillingness of respondent to provide information

The questionnaire aided interviewer is usually an imposition on the time of the respondent, who is required to answer questions on his personal habits, values, socio-economic characteristics and so on to an unknown interviewer. In some cases, respondents refuse to spend the time needed to answer the questions and in quite a few cases, refuse to answer question in relation to their income or on very personal subjects. These 'gaps' in the information collection may affect the overall research process.

Influence of the questionning process

As the situation in which the questionnaire asks respondents to relate, routine actions is an artificial one, they may furnish information which is at variance with the truth.

Quite often, respondentary to give answers that they think will please the interviewer. At times, if the true answer to a question shows them in a bad light or is damaging to

the ego, respondents tend to give 'doctored' or manufactured responses. Also some respondents try at the whole process very lightly and try to amuse shock or astonish the interviewer/reader through their responses.

b) Unstructured direct interviews

This method does not involve the use of a formally structured questionnaire. The interviewer is only given general instructions on the type of information sought. He has therefore a considerable degree of freedom in choosing the questions, as well as the wording and the order which may seem most appropriate for a given interview. Generally used in exploratory research studies these interviews are useful in obtaining a clearer understanding of the problem and determining the areas to be investigated.

This technique as also applied for obtaining information on motives for respondents action. When so used, it usually takes the form of a depth interview. Exploring the motive behind a particular purchase the interviewer may continue asking probing questions like "could you explain what you meant by this statement ? Why do you feel like that? Do you have any other reason"? Until he has obtained all the information he needs. The constraints in this case are requirements of the problem, time limitations and the willingness as well as ability of respondents to put their motives into words.

The advantage of unstructured direct interview is that since it is free of the restrictions imposed by formalistic questions, the interview may be conducted in a conversational casual mode, which is more conducive to information gathering. Moreover the vocabulary can be adjusted to ensure rapport and understanding. The flexibility and informality of the direct unstructured interview often provides access to information not usually obtained in structured questionnaire. The possibility that respondents will knowingly or unknowingly furnish wrong responses is however still there.

The success to the unstructured interview depends upon the skill of the interviewer in formulating and asking questions. The use of highly skilled and competent interviewers adds to the cost per interview. Another factor contributing to costs is that unstructured interviews are much longer than structured ones, and the varied order and form of questions makes editing and tabulating much more complicated.

c) Structured and Unstructured indirect interviews

In order to overcome the nonresponse or inaccurate response problem associated with direct questionnaires and unstructured interviews, some techniques have been developed to elicit information in an indirect fashion. Applying the projective techniques developed in the field of psychology, the respondents are given to describe a non personal ambigous situation. The premise is that the description by the respondents would involve a projection of their own personalities, values, motives, needs and desires. The most common among these techniques are word association tests, sentence completion tests and thermatic appreception (interpretation of pictorial representation). These techniques have been mostly used in the study of buying motivations and consumption patterns of consumer products which are very similar in performance, price and quality, so that the consumer has little basis of differentiating them except brand and company image. Examples are soaps and detergents, food products, cigarettes etc.

Since predesigned set of words, statements or pictoral representation are used to elicit respondents descriptions, all indirect interviews involve some structuring. The degree however varies from highly structured ones to relatively unstructured interviews where the interviewer has a high degree of latitude in questioning the respondent to get fuller responses. Some of the commonly used indirect interview techniques are described below to help you to understand these tools of data collection.

Word association tests

The word association tests involves the presentation of a series of stimulus words to a respondent who is asked to quickly supply the word that first comes to his mind after hearing the stimulus word. Presumably the respondent would give the word that he most closely associates with the stimulus used. For example given the name of a particular brand of instant coffee you may respond in terms of your feelings regarding its flavor,

its evaluation vis-a-vis ground coffee, its costliness, its convenience depending upon the dominant association that you have for this product. It helps in identifying respondent perceptions about products and services, new packaging designs, advertising compaigns and so on.

Sentence completion tests

Developed as an extension of the word association test the sentence completion test involves providing an opening phrase of the sentences to the respondent who is then asked to complete it quickly. Trying to identify the motivation of car buyers a marketing research study used the sentence completion test. The phrases provided were:

People who drive a Maruti.....

Professionals usually drive.....

Most of the new cars .....

When I drive fast .....

Family men prefer .....

The responses given by men and women were found to differ. Fuel economy, maneverability and stylish looks were interpreted to be the more important motives for men while safety was for man important to women.

Thematic appreciation test (Picture interpretation tests)

The test involves showing one or more pictures/cartoons to the respondent and asking them to describe the situation or to assume the role of one of the people shown in the picture/cartoon. Usually the picture shows one or more persons in an ambiguous situation and it is expected that the respondent will decipher the situation in accordance with his own personality.

In marketing research situations, use of cartoon involving people and the product under study is frequently made. The descriptions provided by the respondents are then interpreted by skilled interpreters to understand their price and quality association with the product, its valuation in the mind vis-a-vis competing products and soon.

Another variation of this projective technique is the story completion test when the respondents are given the beginning of a situational narrative and are asked to complete it.

Focus groups

Focus groups are the most widely used type of indirect interviews. Here a group of people jointly participate in an unstructured indirect interview. A purposively selected group of 8-12 people who have a common background on similar user experience relating to the problem under study is brought together in a group discussion exercise. The interviewer, who plays the role of a moderator attempts to focus the discussion on the problem areas in a nondirected, relaxed way. The purpose of this exercise is to provide interaction and involvement among the group members doing the interview, so that a spontaneous discussion with expressions of attitudes opinions values user experience and intention of further product use takes place.

Focus groups characterize preliminary and exploratory studies rather than casual research and are used to generate information to facilitate the formulation of hypothesis or settling of premises for further research. it finds application in studies involving examination of new product concept, finalizing creative concepts for advertising generation of information for modification of existing products as well as determining improved ways of distributing products.

Media used in direct and indirect interviews

The interviews may use the personal medium, involving the use of an interviewer or use mail as a medium. Mail surveys are very common and offer the advantage of a lower costs. However, the incidence of nonresponse in our country is so high that it has severely limited the use of this economical medium.

Another medium though not used frequently in India, is the telephone which lends itself to both structured and unstructured interviews. In countries where telephone interviews are exclusively used they are found to be a speedy and inexpensive medium for conducting interviews.

Let us now look at the other category of data collection methods.

Observation method

Observation methods represent another class of data collection methods where without interacting with the respondents or the subjects under study, observations are made and recorded. Sometimes it is less costly and more accurate, if instead of asking the respondents their behaviour is carefully observed. Though past behaviour cannot be observed, the result of such behaviour can be. For example, instead of asking what brand of television do you own ? Is it black or white ? It is remote operated ? It would be simpler to look at the set and observe its usage.

Observation may be used as the only means of data collection or in combination with other means. It is quite commonly used in exploratory and monitoring researcher designs. For example, if the newly opened Green Supermarket in Delhi wanted to know if its prices of commodities were competitive, the only way to get the information is to regularly monitor, through observation, the prices of these commodities in other supermarkets and comparable stores.

In certain circumstances, though alternative modes of data collection are available, observation may actually be the preferable method of information collection, because of cost or accuracy consideration. In the section dealing with limitation of survey we have discussed inability or unwillingness of respondents to give information accurately. Observation may become the more suitable mode of data collection in such a situation. Studies on brand buying behaviour have shown that purchase/possession of the reputed or prestige brand is overstated while availing of discounts is understated by the respondents. In such cases observations of buying patterns may give better results.

Observation may be carried out manually or mechanically. In countries like UK, USA and Canada devices like the hidden motion picture cameras are used to observe consumer behaviour in the natural situation. The audiometer is another gadget, (described as the electromechanical equivalent of the respondent diary) used for recording day to day behaviour. It is installed in the televisions of a selected panel of people and automatically records the times the set in turned on or as well as the stations tuned in. The recordings collected are used to get total average and share of audience for the different channels and commercials. These devices permit a complete and accurate recording. In India, some Jewellery stores make use of short circuit cameras.

Manually, the observation method involves the use of trained observer who make and record direct observations of respondent behaviour. Information pertaining to the following questions may be collected by direct observations:

a) Who actually is the buyer for the product in question ?

b) To what extent do they seem to have predetermined idea of buying this product?

c) Did they appear to be influenced by the person accompanying them?

d) How many of them checked the price and/or compared it with that of competitors?

e) How many of them studied the package before buying?

Direct observation discloses what was purchased rather than why it was purchased. As motives cannot be observed they must be inferred and these inferences can seldom be substantiated. Another limitation of manual observation is the observers bias — observer may selectively record actions that seem meaningful to them and ignore others. This has implications for the objectivity of the research.

4.8 SOURCES OF ERROR IN PRIMARY DATA COLLECTION

You are familiar with the sources or error associated with secondary data. The primary data collection methods are also subject to three important types of errors. These are sampling error, non-response error and response error.

Sampling error as the name implies is inherent in the procedure of sample chosen and results in the sample becoming non-representative of the population. For example, in order to study the patterns of cigarette consumption among Indian males if you chose a sample of college student in a metropolitan city, this sample would not representative of the population of males in India. The study that you conduct on this sample, no matter which tool of data collection you use, would not be valid because it suffers from sampling error. The range of sampling error however can be controlled by changing the characteristics of sample drawn. Moreover, the extent of the sampling error can be measured if we take a probability sample. More about sampling error has been discussed in the next unit on sampling.

A non-response error occurs when a unit (unit here may be an individual, a family or an establishment) included in the sample, cannot or has not been reached. For example, in a sample of housewives from a particular city area, if a number of them happen to be away everytime the interviewer chooses to come, non-response is likely to occur. Incidence of non-response error as already noted is very high in mail interviews as respondents simply ignore the questionnaire received by them.

In most direct structured interviews i.e. surveys involving use of questionnaires, non-response bias is a sizeable error. It may affect completeness as well as objectivity in data collection as families who cannot be reached after certain attempts during the day may be significantly different from those which can be easily contacted. The non-response error is a serious matter because the direction of the error is generally unknown. One can assume that the non response respondents would each have responded in a given way and therefore can determine the maximum error due to non-response but it is difficult to measure the magnitude of the error. One simple way of minimising this error would be to fix up an appointment before the interview but specially in a country like ours where a large number of respondents do not have access to the telephone, this may not be very practicable.

Response error occurs when the value of the reported variable differs from the actual value of that variable. We would here include errors of both communication and observation. We have already talked about two reasons for response error i.e. inability of the respondent to give accurate information or their unwillingness to give accurate information because of time factor, prestige factor and invasion of primary factor. Let us now discuss the sources of response error related to the investigator and the tools used by him.

Inaccurate information due to the investigator

The most common cause of this type of inaccuracy is cheating by the interviewer. There are a number of way in which interviewers deliberately obtain inaccurate information and supply it. If the questionnaire happens to contain a question that the investigator finds embarrassing to ask, he may decide to supply his own answer or supply an inference on what the respondents answer would have been. In extreme cases reports of interviewees' without ever having contacted the interviews have been discovered to be submitted. Another in between situation that is found to exist is that interviewers get their own friends and associates to fill in the questionnaire or respond to a direct interview and list the responses in the names of the people listed in the sample, thus vitating the entire sampling exercise.

Experienced marketing research agencies feel that like other petty forms of cheating, interviewer cheating can only be controlled to lower its incidence, it can never be eliminated completely. Care in selection, training and supervision of interviewers can

and does help in controlling the incidence of cheating. In addition, certain control procedures like cross checking of small samples of respondents and use to cheater question which disclose the fabricated answers with a fairly high success rate are employed to minimise the incidence of interviewer generated inaccuracy.

Ambiguity

Ambiguity which is defined as errors made in interpreting behaviour or words spoken or written is source of error which occurs in both, communication and observation methods of data collection. All languages are capable of being ambiguous as the person transmitting information and the person receiving them are two different people and the interpretation of the question/behaviour may differ from one person to another.

4.9 SUMMARY

In this unit we have discussed the main types of data used for marketing research in the context of the different research designs that may be employed. The sources of secondary and primary data have been discussed. The two basic methods of data collection i.e. through communication and through observation have been explained alongwith their various modification. The unit ends with a discussion of the sources of error in data collection.

4.10 SELF-ASSESSMENT QUESTIONS

1) One of the important reasons for the use of surveys is that they can obtain sound information on what people actions in the future will be.

   Do you agree? Give reasons for your answer.

2) Discuss the main sources of primary and secondary data.

3) What is the type of data available from official publications?

4) What are the limitations associated with the use of secondary data?

5) What are the tools of collecting data from respondents?

6) Discuss the important sources of error in both secondary and primary data.

UNIT 5 SAMPLING

Objectives

After reading this unit, you should be able to:

- describe the importance of sampling in marketing research studies
- define various basic sampling concepts
- explain the process of sampling in marketing studies
- discuss various sampling designs and their applications in marketing research
- estimate the sample size required in a marketing study

Structure

5.1 Introduction

5.2 Sampling — Its Importance and Limitations

5.3 Sampling: Basic Concepts

5.4 Steps in the Sampling Process

5.5 Sampling Designs

5.6 Estimation of Sample Size

5.7 Summary

5.8 Self-assessment Questions

5.9 Further Readings

5.1 INTRODUCTION

Sampling plays a vital role in carrying out any marketing research study. There would be hardly any marketing research study which does not involve the use of sampling. It is the backbone of marketing research. A company selling a particular brand of toothpaste and interested in knowing what proportion of households in Delhi use its brand would involve the selection of samples of households. A market potential study aiming at finding the numbers, distribution and socio-economic characteristics of potential customers of a product would definetly involve the selection of some kind of sample. company interested in introducing a new style of packaging for its product would be interested in knowing the reactions of its customers — a study of which would require the selection of some type of sample.

In this unit you will be introduced to various sampling concepts. Census (Complete enumeration), an alternative to sampling would also be discussed. A brief mention of sampling and non-sampling errors will be made. The various probability and non-probability sampling designs as applicable to marketing research will be introduced. An important decision while taking a sample is to know how large a sample should be taken since the choice of sample size involves various elements like time, money accuracy etc. Therefore, the determination of sample size would also be discussed in brief in this unit.

5.2 SAMPLING — ITS IMPORTANCE AND LIMITATIONS

You would hardly come across an individual who is not familiar with sampling. People are used to drawing inferences about a large number of items possessing a particular attribute (usually referred to as a population) based on an analysis of a fraction of such items (called a sample). The following examples would illustrate as to where the analysis of samples can be used to draw conclusions about the population.

1) Your mother might like you to taste a small quantity of egg pudding to form an opinion about the acceptable quality of the entire dish.

2) You might have noticed at the book-stalls that before buying a book or a magazine, people generally flip through its few pages (a sample) to determine whether it is of interest to them.

3) A manufacturer of electric bulbs interested in estimating the average life span of a bulb would put a few bulbs (a sample) for complete use to achieve the objective.

4) You don't have to drink a whole bottle of squash to say that it tastes good or bad — a small amount should be able to do the job.

5) You go to the market to buy a bag of 'wheat flour'. To determine its quality, you take a handful of it and examine its quality to draw conclusions about the quality of the whole bag of wheat flour.

These are just a few examples from everyday life where samples are used to draw inferences about the entire population (universe). An alternative to sampling is complete enumeration (census). Under complete enumeration, we collect data for each and every unit (person, household, factory, shop etc.) belonging to the population, which is the aggregate of all units of a given type under consideration. Sampling has some major advantages over complete enumeration and therefore it is used very often in marketing research studies. Explained below are some of the advantages of sampling over complete enum-

1) The data collecue lurogh saniping dcfintcly involves a lowor as compared to census In a markcrung reseaich projcct aming to study the buyong belwiv-iour or , in Delhi for consumer non-durable items, the cost of interviewing 3000 households would be much less than interviewing all the households in Delhi. Therefore, from a managerial points of view, a sample's economic efficiency is very attractive.

2) A sample saves time. Managers have a time frame in which they are supposed to take decision on the basis of whatever information they can gather during that time period. They cannot wait till the information obtained on the basis of complete enumeration is available. In the above example, if data is to be collected on all the households in Delhi, we would require more time to print additional questionnaires, more interviewers will have to be trained and further the data analysis would require more time.

3) It is not at all essential that the accuracy of information may be enhanced by taking a complete enumeration. Suppose we want to estimate the average price of Campa Cola charged by restaurants in Delhi at a point of time. As we are aware that prices usually don't vary widely over a cross-section, it would be a wastage of time and money to go for complete enumeration. One could always obtain very reliable estimates by taking a representative sample.

4) A sample is better in situations in which measuring of a particular element from a group would destroy the elements or render them useless after examination. Let us consider the example of testing a photographic film. To test the quality of the film, the film has to be exposed which infact destroys it for further use. It is not possible to go for complete enumeration as there would be no film left if all were tested. Therefore, a sample of each batch should be used in a particular production run to judge the quality of the film.

The above discussion should not lead you to conclude that carrying out a census is useless. There are some situations under which a census may be preferable to a sample. A census is desirable when the population is small, the variance in the characteristic being measured is large and the fixed cost of sampling is very high. There are also problems associated with sampling which we would discuss shortly.

Problems Associated with Sampling

Let us consider a hypothetical population of size 10,00,000 households whose monthly food expenditure is of interest to us. If we know the food expenditure of all (10,00,000) households in the universe, we could then compute its arithmetic mean

(usually referred to as population mean). Let us assume that the population mean (μ) is Rs.2,500. However, in reality it may not be possible to go for a census for the reasons mentioned earlier. Therefore, one might have to resort to sampling. Suppose a sample of size 1000 is taken, the analysis of which is used to draw inferences about population parameters. Suppose the sample mean ($\overline{X}$: average monthly food expenditure) is Rs. 2,450. We find that though the sample mean is close to the population mean, it is not identical with the true mean (μ) . Suppose we take another sample of size 1000 and compute $\overline{X}$. Let the sample mean in this case be Rs.2600. Here also we find a similar result, that though the sample mean is close to population mean, it is not identical with the true mean. In fact, it is very rare that the sample means would coincide with the true population mean. Why does this happen? There must be certain problems with which you should get acquainted with. We know that the sample data represents only a fraction of population data. Therefore, the problem is related to how well the sample represents the characteristic of the population of which it is a part. The confidence attached to the sample data is affected by two different types of error called sampling error and non sampling error. If both errors don't exist then in our example, sample mean $\overline{X}$ would exactly coincide with population mean for each sample drawn from the population. The two different types of errors viz. Sampling error and non-sampling errors are discussed below:

Sampling Error

The error which arises due to drawing inferences about population parameter on the bais of observations drawn from a sample (a part of the population) is called sampling error. In other words a sampling error is made while selecting a sample which is not representative of the population. It represents the difference between sample value and true value of population parameters. A sampling error is bound to occur while selecting a sample as it is difficult, if not impossible, for a sample (a small part of the population) to be exactly representative of the population. This occurs no matter how careful the researcher is randomly choosing the sample. Sampling error, therefore, is a result of chance. The sampling error usually decreases with increase in sample size and it is non-existent in a complete enumeration survey.

Non-sampling Error

You might expect that a complete enumeration of all the units in the population would result in data free from errors. However, this is not the case. It is difficult to completely avoid errors of observations or ascertainment. While processing the data, one may commit tabulation errors. The main reasons for the occurance of non-sampling errors are the improper specification of the scope of study, inadequate coverage of population and sample, vague definition of variables involved and wrong methods of data collection. A combination of one or more of following factors could always result into non-sampling errors.

1) Inadequate specification of data which may not be consistent with the objectives of survey.
2) Inaccurate methods of interview.
3) Inaccurate reporting by the respondents.
4) Lack of trained and experienced interviewers.
5) Plain lying by the respondents.
6) Inability to locate proper respondents due to improper instructions or wrong address.
7) Inadequate scrutiny of the basic data.
8) Errors in coding, editing and tabulation of data.
9) Errors in presenting the tabulated results, graphs etc.

The non-sampling error can occur in the case of complete enumeration as well as in sample survey. However, non-sampling error increases with the increase in sample size. Therefore, the size of non-sampling error is much larger in case of complete enumeration than that of sample surveys. This is because in a study involving complete enumeration of the units of the population, we need more interviewers, more staff to supervise these

interviewers and more manpower to convert raw data to computer input. This would involve more training of the interviewers, which would be both, costly and time consuming. Further, as the size of staff becomes larger their quality falls. It also becomes more difficult to control and supervise their activities. All this would result into more errors and less accurate results thereby contributing to increased non-sampling errors.

Activity 1
List out some marketing research studies carried out in your organisation or the organisation you know of where sample surveys were conducted.

.....

.....

.....

.....

.....

Activity 2
List out the advantages of sampling over complete enumeration (relate it to the studies in Activity 1).

.....

.....

.....

.....

Activity 3
It is possible to control sampling error? If so, how?

.....

.....

.....

.....

Activity 4
Make a list of the sources of non-sampling error. What do you think should be done to control them?

.....

.....

.....

.....

5.3 SAMPLING: BASIC CONCEPTS

Element : An element is the unit about which information is collected. This provides the basis for analysis according to a well defined procedure. It is important that these should be well defined and it is possible to identify them physically. In marketing research studies, the households, family, shop, factory etc. might constitute elements. The type of unit considered in any marketing research study would depend upon the objectives of the study. For instance, a shop may be considered as a unit in a retail stores survey whereas a household may be treated as a unit in a family budget enquiry.

Population : Population or universe is the aggregate of all elements possessing certain specified characteristics which the researcher wishes to study and defined prior to the selection of sample. A properly designated population must be defined in terms of elements, sampling units, extent and time. Let us consider the following examples.

1) A survey of consumers might specify the relevant population as:

   a) Element: Males 35-40
   b) Sampling Unit: Males 35-40
   c) Extent: Delhi
   d) Time: March 10-April 30, 1990

2) Another survey of consumers might specify the relevant population as:

   a) Element: Females above 20 years
   b) Sampling Unit: Households, then females over 20 years
   c) Extent: Delhi
   d) Time: April 15-25, 1990

3) For conducting a socio-economic survey in a region, we might specify the relevant population as:

   a) Element : Households
   b) Sampling Unit: Tehsils, Villages; then households
   c) Extent : Amritsar District
   d) Time: 1989

Sampling Units : A sampling unit is that element or elements considered available for selection in some stage of the sampling process. In single stage sampling, sampling units and the elements are same. For instance, in our first example of specifying the population (see example above), both sampling units and elements were "males 35-40". This is an example of direct single stage sampling process as the sample of males 35-40 is selected directly.

In a more complex situation (see example 2 above), we are selecting females over 20 indirectly through a two stage process. First of all, we select a sample of households. Then, within these selected households, we select a sample of females over 20. It is only at second stage (final stage in this case) where elements and sampling units are identical.

The third example is more complex than even the second one. In this case a sample of households may be selected in three stages. First of all a sample of Tehsils is selected. Then a sample of villages is selected from each selected Tehsils after making a list of all the villages in it. Finally a sample of households is selected from each selected village after listing all the households in it. In this example Tehsils are taken as first stage unit, villages as second stage unit and households as the third or the final stage unit. You may note that it is at the final stage that elements and sampling units are identical.

Sampling Frame : It is a list of all sampling units belonging to the population to be studied with their proper identification and available for selection at a stage of sampling process. In fact, the actual sample is drawn from the sampling frame. Therefore, we should ensure that the sampling frame contains all the sampling units of the population under consideration. It should exclude units of any other population. The sampling frame should be upto date and free from errors of omission and duplication of sampling units. In fact in marketing research studies a lot of time and effort is spent on preparing a suitable sampling frame. The examples of sampling frame are telephone directory, a list of registered voters, names of all the students registered with a particular university, a map, an employee list etc.

Usually the availability of sampling frame defines the population. Each stage of sampling process requires its own sampling frame. For instance, in three stage sampling process (see, example 3 under population), we would need three sampling frames. These are (1) a list of tehsils (2) a list of villages under the selected tehsils (3) a list of households within the selected villages.

Although a lot of attention is paid to preparing a suitable sampling frame, most of them are far from being perfect. For example, if we are conducting a study which involves surveying the households of Delhi by telephone, the most obvious sample frame which would come to our mind is the use of telephone directly. However, this may not be an appropriate frame as all the households in Delhi may not have telephone connection for

one reason or the other. Further, some numbers may not be listed in the telephone directory which others are listed may not longer be residents of Delhi. Therefore, the survey would not be free from the inaccuracies of population's frame leading to frame's error. In reality, the sampling frame of most of marketing research studies have one or the other shortcomings.

Study Population : A study population is the aggregation of elements from which the sample is actually drawn. You might recall that previously population was defined "as the aggregate of elements (possessing certain characteristics) prior to the selection of sample". Because of certain unavoidable problems, the actual sample is selected from somewhat different population from the one defined prior to the selection of sample. This is because it is very seldom that every element which satisfies our definition of a population actually has a chance of being selected. Our list may be incomplete as some elements are likely to be omitted from a list of population because of certain reasons such as some people may have unlisted phone numbers, a map may not include a new street. a list of registered voters may be incomplete.

Therefore. the study population is the aggregation of elements from which the sample is actually drawn and it is with reference to this (study) population that the inferences are drawn.

Activity 5

Make a list of few marketing research studies undertaken in your oganisation or an organisation you know of where the elements of pupulation were selected directly (single stage process) and indirectly (two stage or three stage or multistage process).

.....

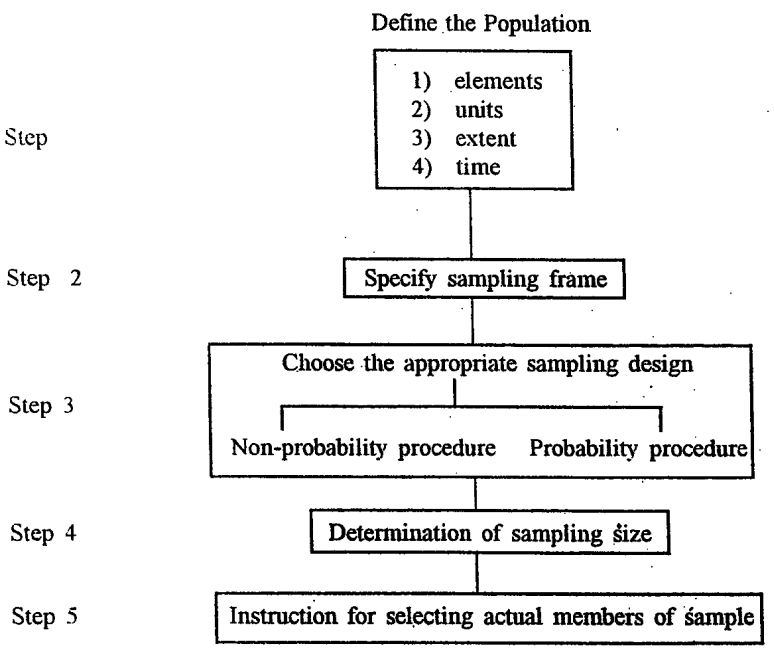
.....

.....

.....

5.4 STEPS IN THE SAMPLING PROCESS

The sampling process consists of five sequential steps. These steps are listed and briefly explained below. An overview of the various phases of sampling is shown in the Figure below.



Step 1 : The primary step is to define the population. It should be defined in terms of (i) the elements, (ii) the sampling units, (iii) the extent, and (iv) the time.

Step 2 : The next step is to specify the sampling frame. It has already been explained in detail in the previous section.

Step 3 : The third step is to choose an appropriate sampling design for selecting the same. The sampling design describes the procedure by which a sample is selected. There are two types of procedures namely non-probability sampling procedures and probability sampling procedures. Under probability sampling procedure, each element has a known chance of being selected in the sample. In the non-probability sampling procedure, there is no known chance of an element of the population being selected in the sample. The selection of an element of population in the sample depends upon judgement of the researcher or of the field interviewer. A detailed discussion on this will be taken up in the next section.

Step 4 : Once we have decided the procedure by which sample would be selected, the next step is to determine how large sample should be taken. A detailed discussion on the determination of size of the sample will find a place in section 5.6.

Step 5 : The interviewers who are to go to field to collect actual data need very clear and accurate instructions as to how to do this job. The sampling planning is therefore not complete until these instructions are prepared and handed over to the investigators.

5.5 SAMPLING DESIGNS

The sampling designs describe the procedure by which sample is selected. There exist two classes of methods by which samples can be selected. They are probability sampling methods and non-probability sampling methods.

Probability Sampling Methods : In probability sampling methods every element of the population has a known chance of being selected. Please note that the term known chance does not mean equal chance. Equal chance probability sampling is a special case of probability sampling, called simple random sampling. In probability sampling methods, there is no chance of arbitrary or biased selection and therefore the laws of probability apply. Therefore, it permits us to measure the sampling error which is the difference between the population value and the sample value.

There are a number of different sampling procedures that fall under probability sampling methods. Some of these methods are listed as under:

1) Simple Random Sampling
2) Systematic Sampling
3) Stratified Sampling
4) Cluster Sampling

We will soon take up the discussion on the above methods.

Non-Probability Sampling Methods : These methods do not provide every item of population any known chance of being selected in the sample. Here there is no attempt to select a representative sample. The elements of samples are selected on the convenience and/or judgement of the researcher or field interviewer. The selection process is subjective. Since the sample is not representative of population, it is not possible to make an estimate of sampling error. Also, we cannot say whether our sample estimates are correct or not.

There are many different sampling methods which fall under the umbrella of non-probability sampling methods. These are listed below:

1) Convenience sampling

2) Judgement sampling

3) Quota sampling

A discussion on these sampling methods will be taken up soon.

It may be worth mentioning that most of the marketing research studies make use of non-probability methods of sampling. Since under these methods, the selection process of these samples are subjective, therefore one should not conclude that the results obtained from it are inferior to what one would obtain by using probability sampling methods. Also, the sample obtained through such methods need not be less representative of the population. For example, a marketing researcher is asked to develop an index of performance of sales force by measuring items such as sales per salesman, number of calls per day, order call ratio, number of customer complaints and so on. Any particular item is included in the list because the market researcher feels that is represents 'performance'. His non-probability method of selection is a better way to achieve a representation of the population than by a random selection of items representing various characteristics of performances.

We would now discuss some of the non-probability sample selection methods that are used in marketing research studies.

Non-Probability Sampling Mehtods

1) **Convenience Sampling :** Under convenience sampling, as the name implies, the samples are selected at the convenience of the researcher or investigator. Here, we have no way of determining the representativeness of the sample. This results into biased estimates. Therefore, it is not possible to make an estimate of sampling error as the difference between sample estimate and population parameter is unknown, both in terms of magnitude and direction. It is therefore suggested that convenience sampling should not be used in both, descriptive and causal studies as it is not possible to make any definitive statements about the results from such a sample.

This method may be quite useful in exploratory designs as a basis for generating hypotheses. The method is also useful in testing of questionnaire etc. at the pretest phase of the study. Convenience sampling is extensively used in marketing studies and otherwise. This would be clear from the following examples.

i) Suppose a marketing research study aims at estimating the proportion of Pan (beetle leave) shops in Delhi which store a particular drink say Maaza. It is decided to take a sample size of 100. What investigator does is to visit 100 pan shops near his place of residence as it is very convenient to him and observe whether a Pan shop stores maaza or not. This definitely is not a representative sample as most pan shops in Delhi had no chance of being selected. It is only those pan shops which were near the residence of the investigator that had a chance of being selected.

ii) The other example where convenience sampling is often used is in test marketing. There might be some cities whose demographic make ups are approximately the same as national average. While conducting marketing tests for new products, the researcher may take samples of consumers from such cities and obtain consumer evaluations about these products as these are supposed to represent "national" tastes.

iii) A ball pen manufacturing company is interested in knowing the opinions about the ball pen (like smooth flow of ink, resistance to breakage of the cover etc.) it is presently manufacturing with a view to modify it to suit customers needs. The job is given to a marketing researcher who visits a college near his place of residence and asks a few students (a convenient sample) their opinion about the ball pen in question.

iv) As another example, a researcher might visit a few shops to observe what brand of vegetable oil people are buying so as to make inference about the share of a particular brand he is interested in.

2) **Judgement Sampling :** Judgement sampling is also called purposive sampling. Under this sampling procedure, a researcher deliberately or purposively draws a sample from the population which he thinks is a representative of the population. Needless to mention, all members of the population are not given chance to be selected in the sample. The personal bias of the investigator has a great chance of entering the sample and if the investigator chooses a sample to give results which favours his view point, the entire sutdy may be vitiated.

However, if personal biases are avoided. then the relevant experience and the acquaintance of the investigator with the population may help to choose a relatively representative sample from the population. It is not possible to make an estimate of sampling error as we cannot determine how precise our sample estimates are.

Judgement sampling is used in a number of cases, some of which are mentioned below:

i) Suppose we have a panel of experts to decide about the launching of a new product in the next year. If for some reason or the other, a member drops out from the panel, the chairman of the panel may suggest the name of another person whom he thinks has the same expertise and experience to be a member of the said panel. This new member was chosen delibertely — a case of Judgment sampling.

ii) The method could be used in a study involving the performance of salesmen. The salesmen could be grouped into top-grade and low-grade performer according to certain specified qualities. Having done so. the salesmanager may indicate who in his opinion would fall into which category. Needless to mention, this is a biased method. However. in the absence of any objective data. one might have to resort to this type of sampling.

3) **Quota Sampling :** This is one of the most commonly used sampling method in marketing research studies. Here the sample is selected on the basis of certain basic parameters such as age, sex, income and occupation that describe the nature of a population so as to make it representative of the population. The investigators or field workers are instructed to choose a sample that conforms to these parameters. The field workers are assigned quotas of the numbers of units satisfying the required characteristics on which data should be collected. However. before collecting data on these units the investigators are supposed to verify that the units qualify these characteristics.

Suppose we are conducting a survey to study the buying behaviour of a product and it is believed that the buying behaviour is greatly influenced by the income level of the consumers. We assume that it is possible to divide our population into three income strata such as high income group, middle income group and low income group. Further. it is known that 20% of the population is in high income group. 35% in the middle income group and 45% in the low income group. Suppose it is decided to select a sample of size 200 from the population. Therefore, samples of size 40. 70 and 90 should come from high income, middle income and low income groups respectively. Now the various field workers are assigned quotas to select the sample from each group in such a way that a total sample of 200 is selected in the same proportion as mentioned above. For example, the first field worker may be assigned a quota of 10 consumers from the high income group, 25 from the middle income group and 40 from the low income group. Similarly the 2nd field worker may be given a different quota and so on such that a total sample of 200 is obtained in the same proportion as discussed earlier.

The above example was a very simple one. however. suppose we are told further that the buying behaviour is not only influenced by his income but also by his age (categorised as 45 and above or below 45). With this additional character, suppose the distribution of population (universe) is as follows:

Table 1 : Distribution of Population (%)

Income / Age	High Income	Middle Income	Low Income	Total
45 and above	12	10	30	52
below 45	8	25	15	48
Total	20	35	45	100

The above table indicates that in this universe there are 12% of people in the high income group and fall in the age group of 45 and above, there are 25% people in middle income group and below the age of 45 and so on. Suppose, it is decided to take a sample of size 200. Therefore, the distribution of the sample conforming to these two parametes (in the same proportion as population) would be as follows:

Table 2 : Distribution of Sample (Numbers)

Income / Age	High Income	Middle Income	Low Income	Total
45 and above	24	20	60	104
below 45	16	50	30	96
Total	40	70	90	200

The above table shows that a sample of 30 should be taken from the population with low income and below the age of 45. Similarly a sample of 20 should be taken from the population with middle income and having age of 45 and above, and so on.

Now, having decided the size of sample falling under each of the six cells ("high income and below 45", "middle income with 45 and above" and so on), we fix the quotas for each of the field worker to collect data conforming to the above norms so as to obtain a total sample of size 200.

At the outset, the Quota sampling procedure might look similar to stratified sampling (to be discussed under probability sampling designs). However, there is a difference between the two. Under stratified sampling, the field worker selects a random sample from each cell of the population, whereas under Quota sampling the selection of sample is not random. It is left to the judgement of the field worker.

The Quota sampling method has some weaknesses. These are listed below:

i) It is usually difficult to obtain an accurate and up to date proportion of respondents assigned to each cell.

ii) As the number of parameters (control characteristics) associated with the objectives of the study become large, the total number of cells increase. This makes the task of field staff difficult as it may not be easy to get a desired respondent.

iii) It is very important that all of the proper parameters (control characteristics) related to the study in question must be incorporated while taking sample. The results of the study could be misleading if any relevant parameter is omitted for one reason or the other.

iv) The field workers might like to visit those areas where the chances of the availability of a respondent satisfying certain desired parameters is very high. Further, the field workers might avoid certain respondents who look unfriendly and live in houses which may not be of good appearance. These factors are likely to make the findings of the study less reliable.

The Quota sampling method has some advantages too. The method has a lower cost and field workers have a free hand to select respondents for each cell to fill their quota. The samples, if selected with care would result into more definitive findings.

Probability Sampling Methods

1) **Simple Random Sampling :** Under this sampling design, each member of the population has known and equal probability of being included in the sample. For details on how to draw samples using this sampling procedure, please review unit 13 of MS-8 course. Simple random sampling is not widely used in marketing research because of the following reasons.

i) In consumer research studies, we usually select individuals, households, shops or areas as the sampling units. It may not be easy to prepare a sampling frame as it is very difficult to get lists of households, individuals and shops, although areas may be completely represented through maps.

ii) We know that an industry comprises of various firms of different sizes. If one wants to study some aspects of an industry, one might like to choose a sampling design where there is a higher probability of a larger firm being selected. **If that is the case**, the very concept of simple random sampling becomes inapplicable in such situations.

The simple random sampling has some applications in Industrial Marketing where generally purchasing agents or companies or areas are the sampling units which are usually not very big in number. Therefore, it becomes easy to prepare a sampling frame thus facilitating the use of simple random sampling.

2) **Systematic Sampling :** The mechanics of taking a systematic sample are very simple. Suppose the population consists of ordered N units (numbered from 1 to N) and a sample of size n is selected from the population in such a way that $\frac{N}{n} = K$ (rounded to the nearest integer). Here K is called a sample interval. Systematic sampling then consists in selecting a number at random between 1 and K (both inclusive) and then selecting every subsequent Kth unit till a sample of size n is obtained.

To make the above more clear, let us assume that we have an ordered population of size N=500. Suppose it is decided to take a sample a size n=50. Therefore, our sampling interval would be $\frac{N}{n} = \frac{500}{50} = 10$. We then select a number at random between 1 to 10 (both inclusive). Suppose it turns out to be 6. Then our sample units would be 6,16, 26, 36 and so on.

Systematic sampling is a case of mixed sampling where both probabilistic and non-probabilistic methods of choosing a sample are used. This is because the first unit of the sample is selected at random between numbers 1 and K (probabilistic method) and then the rests of the units of the sample are fixed by the choice of the first member (non-probabilistic method).

It is very likely that systematic sampling would result into more representative sample than simple random sampling. In systematic sampling the elements of the population are ordered in a particular fashion. Suppose we want to estimate the sales of all the retail stores in Delhi. Under a simple random sampling, if we draw a random sample of size n, it is very likely that most of the sampled stores might turn out to be low sales volume store. However, in systematic sampling we order these retail stores according to ascending or descending order of sales, therefore, a systematic sample would definitely contain some low volume and high volume retail stores. Thus, a systematic sample is likely to be more representative than a sample random sample.

A systematic sample might also reduce the representativeness of the sample. This could happen if items are ordered in such a way to produce a cyclical pattern. Suppose a supermarket is interested in estimating average daily sales by using a sample of certain daily sales. Suppose a sampling interval of seven days is chosen. The sample, would therefore, result in recording of the sales on the same day of the week. This would not reflect day-of-the week variations in sales. For example, a sample of Sundays would generally overstate sales and a sample of Wednesdays and understate sales.

3) **Stratified Sampling :** In stratified sampling, the entire population is divided into various mutually exclusive and collectively exhaustive strata (groups). By mutually exclusive it is meant that if an element of a group belongs to one strata, then it doesn't belong to any other strata. By collectively exhaustive we mean that all the elements of various strata put together completely cover all the elements of the population. The groups (strata) are created on the basis of a variable (criteria) known to be correlated with the variable under study. The possible criteria for stratification of a population could be income of the individuals, age, sex, frequency of a purchase of a product, size of the household, size of the retail store, region of the country and so on. The stratification is also possible on the basis of more than one variable. This, of course, increases the number of stratum. The cost of stratification may come as a constraint in increasing the number of groups (stratum). A variable which is considered to be good in stratification of one population may not be so in the case of other. However, one thing should be kept in mind that stratification should be done in such a way so as to minimise the variability among sampling unit within strata (more homogeneous) and maximise the variability among strata (more hetrogeneous).

Once the population has been divided into various strata, separate simple random sample of various sizes are selected from each stratum. There might be cases, where systematic or any other type of simple random samples may be selected from each stratum and the resulting design may still be called stratified sampling. The selection of the size of sample from each stratum can be done on either proportionate or disproportionate basis. In proportionate stratified sampling, the number of members selected from each stratum is proportional to its share of the total population. However, in case of disproportionate stratified sampling, the number of members selected from each stratum is not proportional to its share of total population. The choice of proportionate or disproportionate sampling method among strata depends upon whether the variances in each group (stratum) are equal or not. If the variance of each stratum is almost equal, one should go for proportionate stratified sampling. If variances are not equal, a large sample should be taken from the stratum with large variance.

Stratified sampling is moderately used in marketing research. Suppose we are interested in estimating the retail sales of Kwality (a brand name) tea in Delhi. Before using stratified sampling method in its estimation, we may ask the question "What factors account for the variations in the retail sales of Tea?" We may possibly get the following two answers.

i) Size of the store-whether it is large, medium or small, Large stores are supposed to sell more tea than small stores.

ii) Day of the week, the sales is supposed to be more during week ends than during the week days.

We have now three types of stores and two types of days of the week. So the total number of stratum (Groups) on which the total population can be divided is six. They are "large store and week days, medium store and weekend and so on". Once we are able to make six stratum, the sample from each stratum can be selected either according to proportionate stratified sampling scheme or disproportionate stratified sampling scheme depending upon the variability in each stratum. One can similarly think of many other marketing research studies where stratified sampling method could be used.

4) **Cluster Sampling :** If we divide all the elements of the population into suitable clusters and select few clusters randomly and all the elements of the selected clusters are used, then this method of sampling is called cluster sampling. This method of collecting data is cheaper since collection of data for nearby units is easier, faster and more convenient than collecting data over units scattered over a region. For instance, it would not only be cheaper but also convenient to collect data on all households in a sample of few villages (clusters) than to survey a sample of the same number of households selected randomly from a list of all households.

The criteria for dividing the population into mutually exclusive and collectively exhaustive clusters is that the elements in the clusters should be as heterogeneous as possible and elements between cluster should be as homogeneous as possible.

To make the concept of cluster sampling very clear let us consider a hypothetical population consisting of 24 elements divided into four equal-sized clusters as shown below:

Cluster	Population element number
Cluster 1	1, 2, 3, 4, 5, 6
Cluster 2	7, 8, 9, 10, 11,12
Cluster 3	13, 14, 15, 16, 17, 18
Cluster 4	19, 20, 21, 22, 23, 24.

Suppose we wanted to select a probability sample of size 12. We could do it in two ways. We could either take a simple random sample of size 12 from the population or alternatively we could select at random two clusters out of four clusters and use all the elements of these selected clusters. In both simple random sampling and cluster sampling, the sample size as a fraction of population size is same. However, all possible combination of elements are not equally likely in cluster sampling. Many of the combinations would be impossible.

5) Area Sampling

In a marketing research study involving sampling of population which may be grouped according to geographical areas (blocks), Census tracts. Communities. constituencies etc., another version of cluster sampling namely Area Sampling is used.

The entire area is divided into various clusters. The cluster may or may not be of equal size. Below we will discuss a sampling scheme where sampling is done by taking into account the size of the cluster. This type of design is called probability proportional to size sampling.

Probability Proportional to Size : This sampling design is used when we have to sample cluster of varying sizes. Suppose there is a town divided into seven blocks (clusters) of varying sizes. Assume that in one of the blocks the size of the household is 120 whereas in the other block the size of the household is 80. The block having 120 size household should have greater probability of being selected than the one with a size of 80. In fact the probability in the first case should be 1.5 times the probability in the second case. This concept is used in this sampling design so that a greater weightage is given to the cluster with higher size.

Let us consider the data given in the following table.

Block No.	Number of household	Cumulative No. of households	Associated random numbers
1.	120	120	001-120
2.	80	200	121-200
3.	210	410	201-410
4.	190	600	411-600
5.	350	950	601-950
6.	150	1100	951-1100
7.	400	1500	1101-1500
	1500		

In the above table a small area has been divided into seven blocks. It is evident from the above table that the number of households in each block are varying. If a sample of size 30 is to be selected from a population of 1500, each household should have a probability of 30 + 1500 = .02 being selected in the sample. We have identified a total of 1500 household in seven blocks (clusters). We will assign a number from 1 to 1500 for each household. This is shown in the last coulmn of the above table. Suppose we have to

select at random three clusters. We will look at the four digited random number from the random number tables with numbers 0001 to 1500. We select three numbers. Suppose the numbers are 476, 1253 & 129. This means cluster numbering 4. 7 and 2 are selected. Now we would select a sample of size 10 from each of the clusters. The probability of selecting the required household are as follows:

Probability of household in block A = Number of blocks to be chosen X block probability X within block household probability

For household in block 2 = $3 \times \frac{80}{1500} \times \frac{10}{80}$ = .02

For household in block 4 = $? \times \frac{210}{1500} \times \frac{10}{210}$ = .02

For household in block 7 = $3 \times \frac{400}{1500} \times \frac{10}{400}$ = .02

We find that irrespective of the block size. the probability of selecting a household equals .02 This is the same as required by the over all sampling design.

We note that in the probability proportional to size (PPS) design the blocks with larger size are given more weightage selected in the sample. It can be shown that the efficiency of the estimate increases by this procedure as compared to when all the blocks (cluster) have equal probability of being selected.

Activity 6

Make a list of sonic of the marketing research studies where some of the non-probalitity simpling designs and probalility sampling designs could be used. Also justify the choice .of a particular sampling design you have selected for a study

.....

.....

.....

.....

Activity 7

Why higher weightage is given to cluster with higher size?

.....

.....

.....

.....

.....

.....

5.6 ESTIMATION OF SAMPLE SIZE

We will discuss below two procedures for estimating sample sizes. The first one is called confidence interval approach which is based on the idea of constructing confidence intervals around sample means or proportions. The second approach is called the hypothesis-testing approach and makes use of both type I and type II errors. The students are advised to revise units 14 and 15 of MS-8 course before reading this section. Let us discuss below the two approaches.

1) The Confidence-Interval Approach

In unit 14 of MS-8. we have discussed the procedure for constructing confidence

intervals for population mean and population proportion. While doing so, we used the following notations:

μ = population mean

$\bar{x}$ = sample mean

σ = standard deviation of the population

s = standard deviation of the sample

$\bar{p}$ = sample proportion

p = population proportion

n = number of items in the sample (sample size)

Suppose we want to use the mean of a random sample to estimate the mean of a population and we want to be able to assert with a $(1-\alpha)$ % confidence that the allowable error of this estimate is e. The question is how large a sample need to be taken to achieve this?

We know that $Z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$

Where Z is a standard normal variate and n is the required sample size. The quantity $\bar{x} - \mu$ represents allowable error e.

Therefore, we have

$$Z = \frac{e}{\dfrac{\sigma}{\sqrt{n}}}$$

$$Z = \frac{\sqrt{n}\, e}{\sigma}$$

$$\therefore n = \frac{(Z\sigma)^2}{e^2} \qquad \ldots (1)$$

The value of Z, for a given confidence level can be read from standard normal table. For a 95% confidence level, the standard normal table indicates that the Z value that allows a 0.025 probability that the population mean would fall outside one end of the interval is Z=1.96. Since we are allowing a total probability of 0.05 that the population mean will lie outside either end of the intervals, Z=1.96 is the correct value for a 95% confidence level. Similarly Z=2.58 is the correct value for a 99% confidence level.

Let us see how can we use formula (1) in estimating sample size for the following problem.

In a study of television viewing habits, it is desired to estimate the average number of hours that teenagers spend watching per week. Suppose the population standard deviation is given to be 192 minutes. How large a sample would be required if one wants to be able to assert with 95% confidence that the sample mean would differ from the population mean by at most 24 minutes?

Here e = 24

σ = 192

Z = 1.96 (for a 95% confidence level)

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

$$= \frac{(192 \times 1.96)^2}{24^2}$$

= 245.86

= 246 (rounded to the next integer)

Similarly if we want to use sample proportion ($\bar{p}$) to estimate the population proportion (p) and we want to be able to assert with a $(1-\alpha)\%$ confidence that the allowable error of this estimate is e. Again we ask the question: How large a sample is required to achieve this?

We know that

$$Z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Where Z is a standard normal variate and n is the required sample size. The quantity $\bar{p} - p$ represents allowable error e.

Therefore, we have

$$Z = \frac{e}{\sqrt{\frac{p(1-p)}{n}}}$$

$$Z = \frac{\sqrt{n}\, e}{\sqrt{p\,(1-p)}}$$

$$\therefore n = \frac{p\,(1-p)Z^2}{e^2} \qquad ..(2)$$

The population p may be unknown. Therefore, in the above formula (2), we substitute the maximum possible value of p (1–p) which can be shown to be equal to $\frac{1}{4}$ and occurs when $p = \frac{1}{2}$. Therefore, in formula (2), we substitute $p\,(1-p) = \frac{1}{4}$ and this would increase our sample size and would ensure that the allowable error lies within the prescribed limit and at given confidence level. Therefore, the formula (2) can be rewritten as:

$$n = \frac{1}{4}\,\frac{Z^2}{e^2} \qquad .(3$$

Let us explain its use with the help of following example:

An automobile insurance company wants to estimate from a sample what proportion of its thousands of policy holders intend to buy a new car within the next one year. How large a sample is required to be able to assert with 95% confidence that the sample proportion and true proportion will differ by less than 0.03?

Here $e = 0.03$

$Z = 1.96$ (for a 95% confidence level)

$$\text{(The Sample Size) } n = \frac{1}{4}\,\frac{Z^2}{e^2}$$

$$= \frac{1}{4}\,\frac{(1.96)^2}{(0.03)^2}$$

$$= 1067.11$$

$$= 1068 \text{ (rounded to next integer)}$$

2) The Hypothesis-Testing Approach

The estimation of sample size based on the hypothesis-testing approach makes use of Type I error and Type II error. Type I error occurs when we reject H_0: the null hypothesis, when it is true. Type II error occurs when we accept H_o when it is false. We will use α to denote the probability of committing a Type I error whereas β is used to denote the probability of making a Type II error. [For details please revise unit 15 of MS-8 course]

We will discuss the estimation of sample size with the help of a numerical example. Let us first of all consider the case of means and then that of proportions.

The Case of Means

A large retailer wants to determine whether the mean income of families living within two miles of a proposed building site exceeds Rs. 6500. It is given that the population standard deviation is Rs. 500. How large a sample will be required if the probability of a Type I error is to be 0.05 and the probability of a Type II error is to be 0.01 when mean income of population is Rs. 6900? To work out this problem, we may note the following:

i) Here, our null (H_0) and alternative (H_1) hypotheses to be tested in terms of population means, μ_0 and μ_1 respectively are as follows:

$H_0: \mu_0 = 6500$
$H_1: \mu_1 = 6900$

ii) We are given

α (probability of Type I error) = 0.05
β (probability of Type II error) = 0.01

iii) For a one tailed test the Z values for the 0.05 and 0.01 risks are found from standard normal tables to be $Z_\alpha = 1.64$ and $Z_\beta = 2.33$

iv) The population standard deviation (σ) = 500. Now we have to calculate the sample size that will meet the α and β requirements. For that, we need to solve the following two simultaneous equations :

$$\text{Critical Value} = \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\text{Critical Value} = \mu_1 - Z_\beta \frac{\sigma}{\sqrt{n}}$$

By equating these two equations and solving n, we get

$$n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_1 - \mu_0)^2}$$

For our example $Z_\alpha = 1.64$, $Z_\beta = 2.33$, $\sigma = 500$, $\mu_0 = 6500$ and $\mu_1 = 7000$

Therefore, substituting these values in the above formula we get :

$$n = \frac{(1.64 + 2.33)^2 \times (500)^2}{(6900 - 6500)^2}$$

$$= \frac{(15.7609)\,(250000)}{160000}$$

$$= 24.626$$

$$\cong 25$$

The Case of Proportion : To discuss the case of estimation of sample size, let us consider the following example:

It has been claimed that 40 per cent of all shoppers can identify a highly advertised trade mark. How large a sample will be required if the probability of a type I error is to be 0.05 and the probability of a type II error is to be 0.01 when the population proportion is 41%?

We set up the hypotheses

$H_0 : p_0 = 0.40$
$H_1 : p_1 = 0.41$

The value of

$Z_\alpha = Z_{0.05} = 1.64$
$Z_\beta = Z_{0.01} = 2.33$

In order to estimate the sample size, we use the following formula

$$n = \left[\frac{Z_\alpha \sqrt{p_0(1-p_0)} + Z_\beta \sqrt{p_1(1-p_1)}}{p_1 - p_0} \right]^2$$

$$= \left[\frac{1.64 \sqrt{0.4(0.6)} + 2.33 \sqrt{0.41(0.59)}}{0.01} \right]$$

$$= \left[\frac{0.803 + 0.564}{0.01} \right]^2$$

$$= \left[\frac{1.367}{0.01} \right]^2$$

$$= 18686.89$$

$$\cong 18687$$

After having decided upon the size of the sample, the field workers should be given clear and accurate instruction for collecting data from the field.

Activity 8

Suppose we want to estimate what proportion of gift items purchased at a department store are returned for a refund, and we want to assert it with 95% confidence that the error of our estimate will be less than 0.05. How large a sample should be taken ?

.....

.....

.....

.....

.....

Activity 9

The personnel director of a manufacturing company wants to estimate the average mechanical aptitude (as measured by a certain test) of a large group of employees, and she wants this estimate to be in error by at most 2.0 with 99% confidence level. If it is presumed from experience that population standard deviation $\sigma = 15.0$ for this test. what should the sample size be?

.....

.....

.....

.....

Activity 10

Suppose we want to test the null hypothesis μ = Rs. 400 against the alternative $\mu \neq$ Rs. 400 for a population whose standard deviation is Rs. 12. If this hypothesis is true, we

want to be 95% confident of accepting it, and if the true mean is Rs. 405, we want to be 90% sure of rejecting the hypothesis. What is the required sample size?

.....

.....

.....

.....

Activity 11

If one sets up the hypotheses

$H_0 : p_1 = 0.4$
$H_1 : p_1 = 0.5$

With a risk of 0.1 and a beta risk of 0.05 if $p_1 = 0.7$, what is the appropriate sample size under the hypothesis testing approach?

5.7 SUMMARY

In this unit, we have highlighted the importance of sampling in marketing research studies. The advantages of sampling over the complete enumeration are explained. The unit also points out the problems associated with sampling and discusses why sampling estimate may deviate from the population estimate (true value). In this connection, a brief discussion on both sampling and non-sampling error finds a place in this unit.

The various steps in the sampling process are also briefly explained. The two classes of sampling methods namely probability sampling methods and non-probability sampling methods are brought out in this unit. Under non-probability methods — the sampling designs like convenience sampling, judgement sampling and quota sampling have been included. The sampling designs like simple reandom sampling, systematic sampling, stratified sampling and cluster sampling find a place under probability sampling methods. The determination of sample size by (i) the confidence — interval approach, and (ii) the hypothesis — testing approach forms the concluding section of the unit.

5.8 SELF-ASSESSMENT QUESTIONS

1) Compare and contrast quota sampling with stratified sampling.

2) Point out the basic difference between stratified sampling and cluster sampling.

3) List out the possible sources of errors in the following procedures.

   i) A basket of apples is sampled by taking some apples from the top.

   ii) Investigators collecting data for a study on household expenditure in a town conduct house-to-house enquiry of the households selected at random, during the working hours of the day, ignoring those houses from which there is no reply.

   iii) For a study aimed at estimating the daily average sale of a supermarket, an investigator collects sales figures on every Tuesday of the week for a period of six months.

   iv) To get the feedback on the new T.V. serial presently being telecast on Doordarshan, the producer of the serial collects the opinion of the people living in his neighbourhood.

4) Explain in detail the various sampling designs under non-probability sampling method. Also bring out their relevance in marketing research studies.

5) Define stratified random sampling. What stratification criteria would you employ in sampling public opinion on TV serial 'Mahabharat'?

6) It is desired to estimate the mean lifetime of a certain kind of vacuum cleaner. Given that population standard deviation σ = 320 days, how large a sample is needed to be able to assert with a confidence level of 99% that the mean of the sample will differ from the mean of the population by less than 45 days.

7) Suppose we want to estimate what percentage of the drivers are actually exceeding the 50 kmph speed limit on a stretch of road between Moti Nagar and Vikas Puri in New Delhi. How large a sample would be required to be able to assert with a 95% confidence level that the sample proportion and the true proportion will differ by less than 4 per cent?

5.9 FURTHER READINGS

Freund. John E. and Frank J. Williams, "Elementary Business Statistics — The Modern Approach", Prentice Hall International Editions.

Kinnear, Thomas C. and James R. Taylor, "Marketing Research — An Applied Approach" McGraw-Hill International Editions.

Luck, David J. and Ronald S. Rubin, "Marketing Research" Prentice Hall of India Pvt. Ltd..

Green, Paul E. and Donald S. Tull "Research for Marketing Decisions" Prentice Hall of India Pvt. Ltd.

Westfall, Boyd and Stasch, "Marketing Research — Text and Cases" Richard D. Irwin, Inc.

UNIT 6 QUESTIONNAIRE DESIGN AND DEVELOPMENT

Objectives

After going through this unit you should be able to:

- Explain the use of a questionnaire as an instrument of data collection
- Explain the different types of questionnaire
- Design a questionnaire

Structure

6.1 Introduction
6.2 Questionnaire: Format and Administration
6.3 Steps Involved in Designing a Questionnaire
6.4 Constructing a Questionnaire
6.5 Management of Field Work
6.6 Coding and Tabulation
6.7 Ambiguities in the Questionnaire Method
6.8 Summary
6.9 Self-Assessment Questions
6.10 Further Readings

6.1 INTRODUCTION

Questionnaires constitute the most prevalent method of information collection among the communication methods used. Both structured and unstructured questionnaire are in use in marketing research. The unit discusses various issues regarding questionnaire, their types as well as the steps in designing a questionnaire. A sample questionnaire is also included to familiarise you with the structure and design of questionnaire.

6.2 QUESTIONNAIRE: FORMAT AND ADMINISTRATION

A questionnaire is a standardised format of data collection. It is normally used when the data is collected from a large population about their awareness, attitudes, opinions, past and present behaviour. While there is no standard procedure to construct a questionnaire, the problem, situation and experience of the researcher greatly influence the process. The respondents may be asked questions without any formal questionnaire or with a pre-planned properly framed questionnaire. There could be a number of variations in the use of questionnaire method of data collection. They could vary in terms of its 'format' and also in terms of the way it is 'administered'.

Questionnaire Format

Questionnaire format depends upon the amount of structure and diguise required during data collection.

a) Structure

At the time of framing the questionnaire the researcher must appropriately determine the degree of structure to be imposed on the questionnaire. A highly structured questionnaire is one in which the question to be asked and the responses permitted are

explicitly pre-specified. On the other hand in a non structured questionnaire the questions to be asked are kept flexible in their own words and also the respondents are allowed to answer the questions in a manner they like. The response pattern may vary from open-ended to closed-ended. In open-ended question the respondent is free to choose the possible response, where as in the closed ended from, the researcher pre specifies certain options and the respondent is allowed to choose the alternative(s) from the given options. For example, the structure of these two forms of response will be as follows:

Open-ended : What brand of shampoo do you use?

Close-ended : Mention the brand of shampoo you use from the list given below:

( ) Ponds ( ) clinic ( ) Tiara ( ) Palmolive

b) Disguise

Disguised questions is one where purpose is not made obvious to the respondents and is asked in an indirect manner. Non-disguised questions, on the other hand, are ones which are direct and the purpose of asking them is known clearly is the respondents. Disguised questions are used in the conditions when the issues concerned are such that respondents may not give correct answer to direct questions.

Based on the above discussion, questionnaires could be classified into for categories.

i) Structured, non-disguised questionnaire

ii) Structured, disguised questionnaire

iii) Non-structured, non-disguised questionnaire

iv) Non-structured, disguised questionnaire

Structured, non-disguised questionnaires are very popular in marketing research studies. These are more applicable when large sample sizes are there. Non-structured, non-disguised questionnaires, on the other hand, are used when a freehand is to be provided to the respondents so that an in-depth information on the subject could be solicited e.g. in industrial marketing research wherein number of respondents would also be low.

Non-structured, disguised questionnaires are mainly used in 'motivation research'. 'Word Association Test', 'Sentence Completion Test', 'Thematic Appreciation Test', 'Cartoon Test', etc. may be used in this category, Structured, disguised questionnaires are more appropriate where responses are required towards certain sensitive issues like attitude towards aids patients, abortion etc.

Questionnaire Administration:

So far we have discussed about different formats of questionnaires. However, the questionnaire method may also vary depending on the way it is administered. These could be broadly classified into three different categories.

i) 'Personal interview', wherein there is a face to face interaction between interviewer (s) and respondents (s).

ii) 'Telephone survey', in which survey is conducted over phone i.e. unlike personal interview there is only a voice contact.

iii) 'Mail survey', as the name suggests, is conducted through mail and as such there are no interviewers.

6.3 STEPS INVOLVED IN DESIGNINING A QUESTIONNAIRE

Step 1, Preliminary decision

a) Assess what information will be sought after a thorough scanning of secondary sources of data.

a) Determine the target respondent.

Step 2. Decide on the type of questionnaire and method of administration.

Step 3. Evaluation of question content. Before including a question in the schedule, examine whether.

a) This question is really essential.

b) The respondent can understand the question. (i.e. is it too technical, ambiguous, or advanced for the target respondent?)

c) The respondent can answer the question. (Say, the respondents possess sufficient knowledge. As such, it is better not to ask too much of factual data or about past history.)

d) The respondent will answer the question. Specially, if it invades into one's privacy or it requires too much effort to answer. then they usually refuse to cooperate.

Step 4. Check question phrasing.

For Example

a) Do words have ambiguity in meaning?

b) Are there any implied alternatives in the question?

c) Are there some assumptions to be made to answer the question?

d) Will the respondents approach the question from the same frame of reference as designed by the researcher?

Step 5. Determine form of response to each question:

a) The response format may be open or closed ended. In general, the type of response format will depend on the objective of the research, nature of data to be collected and analysis to be performed.

Step 6. Determine sequence of questions.

- Use simple and interesting opening questions;
- Design branching questions with care;
- Put the questions in logical manner;
- Ask for classificatory data at the end.

Step 7. Assess the physical layout of the questionnaire. Naturally, the questionnaire must be printed properly, put in elegant form and facilitate handling.

Step 8. Pre-test the questionnaire.

It is normal practice to pretest a questionnaire on a small number of target respondents. The pretest is done to assess both individual questions and their sequence of response pattern. Accordingly, a researcher must revise questions which cause problems.

While developing a questionnaire, researcher must

a) Use simple words in framing the questions;

b) Avoid ambiguous questions;

c) Avoid implicit alternatives;

d) Avoid questions that require too much memory recall and calculation;

e) Avoid double barreled questions;

f) Ideally, a questionnaire should first secure some basic information to get the respondent's cooperation and gradually try to collect more information about the phenomenon of interest.

g) It is easier to administer a multiple choice response categories that requires one simple tick.

6.4 CONSTRUCTING A QUESTIONNAIRE

Marketing Problem

A tyre manufacturing company is planning to lauch a new tyre for scooters/motorcycles. The company would like to have information about how consumers select a brand of tyre for such vehicles.

Marketing Research Objectives

1) To estimate the prevailing level of brand awareness among consumers and to what degree are the consumers able to differentiate between company name and brand image.

2) To identify the attributes of scooter/motorcycle tyres which are important to consumers and to what extent do these attributes influence the buying behaviour.

3) To study the buying process for scooter/motorcycle tyres.

Information Requirements

1) Top of the mind awareness of manufacturers of scooter/motorcycle tyres.

2) Consumer's perception of brands and the importance they attach to brand names.

3) The extent to which the consumers think the following attributes: company name, tread design/road grip, long life and comfortability are present in the tyres they are using.

4) The degree of importance which the consumers attach to the attributes mentioned above.

5) The buying process involved in the selection of a tyre

6) Consumer attitude towards retreading.

Questionnaire

1) Do you own any scooter or motorcycle? Please tick below.
   Scooter ( ) Motorcycle ( )
   If no, terminate the interview

2) What is the make of your scooter/motorcycle?

3) Which tyres are you using on your two-wheeler at present?

4) Please name some two-wheeler tyre companies that you have heard of.

5) What do you understand by the term "brand" of a tyre?
   a) Company name
   b) Quality of the tyre
   c) Identifying name/number on the tyre
   d) Type (Specification) of the tyre.

6) How important is the brand name of a tyre to you?

1	2	3	4	5
Not so important				Very important

7) Please name some two-wheeler tyre brands that you have heard of.

8) How do you rate your present tyre on the following qualities?
   a) Company name

1	2	3	4	5
Less known				Well known

b) Tread design/Road grip

1	2	3	4	5
Poor				Good

c) Price of the tyre

1	2	3	4	5
Low				High

d) Life of the tyre/Hardness

1	2	3	4	5
Low				High

e) Comfortability

1	2	3	4	5
Uncomfortable				Very Comfortable

9) Are you satisfied with the overall performance of your present tyre?

1	2	3	4	5
Very Satisfied				Very Comfortable

10) Have you ever bought a new two-wheeler tyre?

Yes ( ) No ( )

11) Did you change both the tyres at the same time or one at a time?

Both at the same time ( ) One at a time ( )

12) Is your present tyre of the same make as the previous one?

Yes ( ) No ( )

13) a) If No then what was the make of the previous tyre?
b) What made you change over to the present tyre?

14) How did you decide that your tyres needed replacement?
a) Decided on your own
b) Advised by your mechanic
c) Advised by your friends
d) Advised by your service station man
e) Any other, please specify..

15) How did you buy your tyres?
a) You bought them personally from the nearest shop
b) You bought them personally from a specific dealer
c) You asked your mechanic to buy them
d) You asked your friend to buy them

e) Any other, please specify.....

16) How did you decide the make of the tyre?

a) On your own

b) After consultation with your mechanic

c) After consultation with your friends

d) After consultation with the dealer/shopkeeper

e) Any other, please specify .....

17) What are the factors that you considered when you bought the tyre?

a) Company name

b) Price of the tyre

c) Brand name

d) Past Experience

e) Advertising

f) Physical qualities of the tyre

g) Any other, please specify.....

18) Have you ever retreaded your two-wheeler tyres?

Yes ( ) No ( )

If No, then you did not get them retreaded because

a) It is not safe

b) It does not last long

c) It is not worth the saving

d) If yes, then why?

19) How important are the following qualities of a two-wheeler tyre to you?

a) Good company name

1	2	3	4	
Not so important				Very important

b) Tread design/Road grip

1	2	3	4	5
Not so important				Very important

c) Price of the tyre

1	2	3	4	5
Not so important				Very important

d) Life of the tyre/Hardness

1	2	3	4	5
Not so important				Very important

e) Comfortable ride/softness

1	2	3	4	5
Not so important				Very important

20) If your friend asked you which tyre to buy, which one would your recommend?

21) Kindly tick ( ) the ones you have heard of

	Company	Brand
a)	Dunlop	Nylogrip
b)	Ceat	B-82
c)	Modi Continental	Chetak
d)	Vikrant	ANC
e)	MRF	K-Series
f)	Modistone	Vs-101
g)	Metro	
h)	JK	
i)	Srichakra	

22) Please let us know the following information about yourself

a) Your age group (Years)
   i) 18-25
   ii) 26-35
   iii) 36-50
   iv) Above 50

b) Your family income group (Rs. per month)
   i) Less than 2000
   ii) 2000-3000
   iii) 3000-4000
   iv) 4000-5000
   v) Above 5000

c) Your educational qualifications
   i) High school
   ii) Graduate
   iii) Post Graduate
   iv) Any other qualification, specify

d) What is your occupational status?
   i) Own business
   ii) Engaged in service
   iii) Unemployed
   iv) Student

6.5 MANAGEMENT OF FIELD WORK

Data collection in a marketing research study may require to utilise the services of many interviewers. As most marketing research studies involve extensive tours, the interviewer must have good health. Secondly she/he should have a pleasing personality. Thirdly, given the nature of Indian market, they must have command over local and some regional languages. Fourthly, an interviewer should be able to freely communicate with the respondents.

Before the interviewers are sent out for data collection, the interviewers must be thoroughly briefed about the project and manner in which data are to be collected.

It is also necessary to continually supervise the progress of work made by the

interviewers during the survey. It is some times necessary to visit the field without prior intimation to the interviewer in order to check that the interviewers are following the instructions properly. This sort of checking will enable the researcher to assess whether the interviewer has selected the respondents as per the sampling procedure, whether the interviewer has resorted to some mal-practices and also to see the difficulties faced in data collection stage.

6.6 CODING AND TABULATION

Coding is the technical name used to categorize the data collected. It involves specifying the alternative categories or classes into which responses are to be placed and assigning code numbers to the classes.

In fact, coding closed-ended questions are simple, depending on the response pattern one may design the various codes for different responses. Open-ended questions are, however, very difficult to code. Here no uniform rules exit. The researchers have been found to apply their own discretion in developing the coding procedure.

Tabulation consists of simply counting the number of cases that fall into the various categories. It may take the form of simple or cross tabulation. Simple tabulation involves counting the responses based on one variable.

One may apply various statistical tools like mean, median, mode, standard deviation or variance to condense and interpret the data. To present the data in graphical form one may make use of various forms of statistical graphs like histogram, frequency polygon, ogive etc. A brief description of these tools is given below.

Histogram: It is a form of bar chart where the values of the variable are placed on the X-axis and the frequency of each distinct occurrence is indicated on the Y-axis.

Frequency polygon: Figure obtained from the histogram by joining the midpoints of the bars of the histogram with straight lines.

Ogive: This graph shows the number of cases having a value less than or equal to specified quantity; that is, the cummulative frequency is generated. Afterwards the cummulative frequencies are plotted along the Y- axis against the given combination of values of the variable (or X-axis).

Cross tabulation involves counting the incidence of occurance of various combinations of responses of two or more variables. For example, let two variables be various brands of a product and the different regions. Further assume that there are three brands namely A,B,C; and four geographic regions such as East, South, West and North, suppose sales figure of different brands was cross classified against different regions. Then the data will look like following:

Geographic Region

Brand	**East**	**South**	**West**	**North**
A				
B				
C				

It basically shows the joint occurance of different combination of responses. One may compute correlation coefficient to measure the degree of association between the two-way classification.

Activity 1

Assume you are conducting an exploratory study about buying **behaviour for televisions**.

a) Specify the kind of information that is to be sought.

b) Outline a complete questionnaire for this study

c) Explain how the data obtained by your **questionnaire will be coded and tabulated**

d) Attach a set of instructions **for collecting the data by a field force**

.....

.....

.....

.....

.....

6.7 AMBIGUITIES IN THE QUESTIONNAIRE METHOD

Unambiguous communication in the context of marketing research can be achieved if the question asked and the responses given mean the same thing to the interviewer and the respondent. For example in an actual research situation a company trying to find the desirable attributes of it's to be introduced soya fruit drink pretested a questionnaire. One of the question asked was:

Do you like the soya fruit drink?

It is obvious that an assumption is being made here that the respondent in answering the above question would refer to the taste of the soya fruit drink. The pretesting exercise however revealed that housewives who formed the sample said yes to the above question, not because they liked the taste, but because they liked soya drink in terms of the protein input benefit it would bring to their children. Since the objective of this question was to probe consumer perceptions about product attributes like taste, the question after pretesting was changed to:

Do you like the taste of soya fruit drink?

Some ambiguity may be associated with the different forms of questions used. If the form of question used is open ended, it requires the respondent to express himself more fully as well as individualistically. As a result, this form of questioning provides far greater possibility of ambiguity in interpreting responses. For example in the same study with respect to the soya fruit drink the actual response to an open question is given below:

Question : What suggestion do you have to improve soya fruit drink?

Answer : 'I haven't really given it much thought. I suppose it would be better if it came packed in transparent bottles instead of these pouch packages which you cannot use or be too sure of. But then, bottles break."

How would you interpret this response? Could you categories it as non-response? Or can you presume that she has made a clear suggestion that the drink should be packaged in a glass bottle or that pouch packages are not acceptable at all?

The conclusion that can be drawn is that, being short and simply structured, while the open ended questions allow the least possibility of the questions being ambiguous, they result in the highest possibility of answers being ambiguous.

In close ended multiple choice question form, where several set alternative responses are provided, the commonest source of ambiguity is the difficulty of making the alternatives mutually exclusive. For example:

You prefer this brand because you think:

a) it is reasonably priced

b) it is a tried and tested brand

c) it has a better taste and flavour

d) it is good value for money

e) it is a well established brand.

Alternatives a) and d) as well as c) are not mutually exclusive or both present shades of the same of attribute. To the respondent this ambiguity may be confusing.

The multiple choice form is also prone to another kind of ambiguity. It is presumed here that there would be a single most correct response to a given question.

Related as most survey are to consumer behaviour situation there is hardly ever a single behavioural response which may describe respondent behaviour correctly in all circumstances. To give an example:

Q) When your favourite brand is not available on your preferred shop do you:

a) buy the second preferred brand

b) look for it in other stores

c) wait till stocks arrive

d) buy any other brand.

A respondent who normally would like to shop around in one or two shops before switching to another brand but not if he is short of time, would be confused between alternatives a and b.

An interesting type of error associated with forms of question is the tendency of respondents to chose the alternatives appearing first and last much more frequently than those in the middle. This error called the position bias may be reduced to a certain extent by rotating the order of alternatives for different respondents.

In the multiple choice form of question though the possibility of responses being ambiguous is average, that of questions being ambiguous is fairly high.

Some other possibilities of ambiguity may arise because of the following situation:

a) **Question being too long**

Long questions comprising of complex and compound sentence structures become incomprehensible to respondents. As all words are potential sources of ambiguity, the longer the question, greater will be the possibility of its being misunderstood. Consider the following question which was actually used in a study on distribution network for automobile parts:

Q) Under the new system, do you think spare part dealers would be independent businessmen like applicance dealers and furniture merchants who own their outlets, or they would be employees of the automobile companies?

This sort of question would pose problems of comprehension among most respondents. It could easily be rephrased as:

Q) Under the new system, do you think spare part dealers would be owners of their business or employees of the automobile companies?

b) **Question using vocabulary unfamiliar to be respondent.**

The questions should, as far as possible, consist of words that are a part of the normal vocabulary of the respondent. For example consider the following question.

Q) Do you think the pasteurisation process interferes with the lactogenic balance of milk?

This question may be all right if put to doctors, chemists or biochemists but how many members of the general public would be a match to the vocabulary of this researcher?

c) **Question using words that are ambiguous in context:**

Sometimes the words that are fairly understandable on their own may be used in a way that renders their ambiguous. In the following question:

Q) Do you watch television programmes regularly?

The word 'regularly' may have different meanings for different people. It does not specify whether regularly means the whole day long, seven times a week, five times a week, or certain programmes on every telecast. The question therefore needs to be rephrased in the light of the information sought.

d) **Combined questions**

Poor question construction sometimes results in two questions being asked as one. A question put to housewives is given below as an example:

Q) What do you think is a healthier and economic medium for your cooking, refined oil or unrefined ones?

It is clear that the housewife who thought that one medium was more healthy and the other more economical would not be able to respond logically to this question. The simpler and more effective way to get this information would be to break this question into two, one dealing with nutritional value and the other with economy.

From the above illustrations it must be clear to you that all research designs using communication methods of data collection have potential sources of ambiguity inherent in the communication process. As many safeguards as possible therefore should be used to avoid ambiguous question or confusing responses. Pretesting of question and verification of information collected are simple but effective way of ensuring that at least this error becomes controllable.

6.8 SUMMARY

The unit discussed the planning and execution of questionnaires. Steps involved in constructing questionnaires and rating scales have been explained. The errors associated with measurement have also been mentioned. The unit includes a sample questionnaire to facilitate the understanding of the construction, scaling and formatting of questionnaires.

6.9 SELF-ASSESSMENT QUESTIONS

1) What is meant by a questionnaire?
2) What sort of marketing information can be collected with the help of a questionnaire?
3) What are the merits and demerits of a structured questionnaire?
4) In what sort of marketing studies will you use an unstructured questionnaire?
5) Explain the meaning of open-ended and closed-ended question.

6.10 FURTHER READINGS

Freund, John E. and Frank J. Williams, "Elementary Business Statistics — The Modern Approach" Prentice Hall International Edition.

Kinnear, Thomas C. and James R. Taylor, "Marketing Research — An Applied Approach" McGraw-Hill International Edition.

Luck, David J. and Ronald S. Rubin, "Marketing Research" Prentice-Hall of India Pvt. Ltd.

Green, Paul E. and Donald S. Tull "Research for Marketing Decisions" Prentice Hall of India Pvt. Ltd.

Westfall, Boynd and Stasch, "Marketing Research — Text and Cases" Richard D. Irwin, Inc.

UNIT 7 ATTITUDE MEASUREMENT & SCALING

Objectives

After going through this unit you should be able to:

- define attitude.
- distinguish between various types of measurement scales.
- discuss how the method of analysis of data is dependent on the levels of measurements used in collection of data.
- explain different formats of rating scales.
- differentiate between single item and multiple-items scales.
- construct various types of multiple-items scales.
- describe criteria for good measurement.

Structure

7.1 Introduction
7.2 Attitude
7.3 Levels of Measurement
7.4 Formats of Rating Scales
7.5 Single-item vs. Multiple-Items Scales
7.6 Criteria for Good Measurement
7.7 Summary
7.8 Self-assessment Questions
7.9 Further Readings

7.1 INTRODUCTION

Marketing research generally deals with the behaviour of the consumers. A company which is able to sell its product/services must be viewed by customers as pocessing a favourable attitude towards it. It is therefore very important for the organization to understand and measure the attitudes of its customers towards its products and services. This is because if the customers have an unfavourable attitude or/a poor image about the company, it will not be able to sustain for a long time. It is not only essential for the company to ensure that consumers have a favourable attitude towards its product, but also to anticipate their future preferences. The question therefore arises how to measure attitude towards the products of the company. Measuring attitude is a very difficult task because we cannot measure product or customers but we can measures the opinion/ perceptions of the customers towards the product. In other words, attitude is inferred and not directly observed. A study of consumer behaviour and likes and dislikes of the customers play a very important role in understanding their perceptions and thereby inferring their attitudes towards the product. This may be of help in estimating how much of the product a company will be able to sell in future.

7.2 ATTITUDE

Attitude may be defined as the degree of positive or negative affect associated with some psychological object. It is a pre-disposition of the individuals to evaluate some

object or symbol or aspect of his world in a favourable or unfavourable manner. Attitude comprises of three components.

1. **A cognitive component** — a person's belief or information about the object.
2. **An affective component** — a person's feeling about the object such as "like" or "dislike", "good" or "bad"
3. **A behavioural component** — a person's readiness to respond behaviourally to the object.

The study and measurement of attitudes is important since it is assumed that there is a relationship between attitude and behaviour. The research, however, indicates that such a relationship hold more at aggregate level than at the individual level. Attitude may only be one of the factors influencing behaviour - there could be other factors beside attitude which may be more powerful in influencing behaviour. For example, an individual having a favourable attitude towards a product may not buy it because of economic considerations. For the purpose of marketing decision the attitude behaviour relationship relates to measuring of cognitive and affective components and being able to predict future behaviour.

7.3 LEVELS OF MEASUREMENT

Measurement means the assignments of numbers to objects or persons to represent quantities of their attributes. The assignment of numbers is done according to some rule. The attribute of person could include his income, preference, religion, social class, attitude etc. Similarly we measure a product's speed, colour, size, flavour etc. It is the characteristic of the buyer or the product which is measured and not the person or product itself. There are four types of scales used in marketing research to measure attitude towards a particular product/service. These are as follows:

1. Nominal Scale
2. Ordinal Scale
3. Interval Scale
4. Ratio Scale

1. **Nominal Scale :** In this type of measurement numbers are used to label persons, objects or events. For example, the variable sex may be categorized as male or female. One may assign a number 1 to male and 0 to female. It only helps us to identify that whenever number 1 is used we are talking about males and 0 in case of females. Similarly if the units in the population is to be classified according to religion viz. Hindu, Muslim, Sikh and Christian; one may label Hindu as 101, Muslim as 102, Sikh as 103 and Christian as 104. The numbers, here, have no meaning in the sense that the category which is assigned a higher number is in no way more important (or bigger) than the category which is assigned a lower number.

   The classification to identify objects, events and person is also done on the basis of the letters of english alphabets. Using the example mentioned in the preceding paragraph one could have labeled males as A and females as B. Similar procedure may be used for another type of classification. However, the classification should be done in such a way that groups are mutually exclusive and collectively exhaustive.

   The only mathematical operation performed on nominally measured data is the count in each category. Numbers assigned to represent categories (e.g. 1 for male and 0 for female) cannot be added, subtracted, multiplied or divided. We can only say (if sample comprises of 100 respondents) that there are 70 males and 30 females. Also we can say that the sample comprises of say 40% Hindus, 20% Sikhs, 30% Muslims and 10% Christians. The statistical procedures applicable for nominal scale measurement are Mode and Chi-Square.

2) **Ordinal Scale :** A significant amount of consumer oriented research relies on ordinal measurement. Here numbers, letters or any other symbols are used to rank items. Ordinal scale tells us whether an object or event has more or less of a characteristic than some other object or event. Unfortunately this scale doesnot indicate how much more and how much less we have of the characteristics the objects or events pocess. For example, we may be examining the marks in Marketing for 60 students in a class. Assume that the highest score in the class is 92, the second highest 63 and the third highest 60. We rank the person getting 92 marks as 1, 63 marks as 2 and 60 marks as 3. However the difference between the marks of first and second rank is not the same as the difference in the marks of students obtaining second and third rank.

The common mathematical operations like addition, subtraction, multiplication and division cannot be used with ranked data. However, statistical procedure based on interpretation of "greater than" or "less than" are permissible. The statistical methods applicable with ordinal data are percentiles, median and rank order correlation.

The ordinal scale measurement is of higher level than the nominal scale measurement. i.e. to say the ordinal scale data can also be converted into nominal scale data and thus will pocess all the properties of nominal scale data. As an example assume that 107 consumers are divided according to their income and the classification is as given below:

Monthly Income (Rs.)	No. of Consumers
Less than 5000	20
5001 - 8000	30
8001 - 11000	25
11001 - 13000	20
13001 and above	12
TOTAL	107

A consumer having an income of Rs. 7500 will be ranked higher than the one having income of less than Rs. 5000 (Ordinal scale data). However, the sample of 107 consumers can be classified into 5 income categories to be labeled as A for less than Rs.5000, B for 5001-8000, C for 8001-11,000, D for 11001-13,000 and E for 13001 and above (nominal scale data). A consumer having a monthly income of Rs.12,000 will belong to category D..

3. **Interval Scale :** Interval scale responses are more powerful than ordinal scale responses. They not only pocess the properties of ordinal and nominal scale measurement but also the strength of the equality of differences between ranks. For example, the consumers may be asked the following question.

How important is price to you while buying furniture?

Very important	Important	Neither important nor unimportant	Unimportant	Very unimportant
(1)	(2)	(3)	(4)	(5)

One may note that very important is assigned a number (1), important as (2) and so on and lastly very unimportant is assigned a number (5). What is assumed here is that the respondent is able to reply on a continuum scale and the difference between any two responses can be meaningfully interpreted.

The general mathematical form of interval scale is given by the equation.

$$Y = a + bX$$

Here we are considering an arbitrary zero point or starting point and therefore the division of two responses have no meaningful interpretation. Instead of giving number 1 to very important and 5 to very unimportant in the above example, if the numbering starts from 0 to 4 the picture would look like as shown below in Scales A and B.

	Very important	Important	Neither important nor unimportant	Unimportant	Very unimportant
Scale A	(1)	(2)	(3)	(4)	(5)
Scale B	(0)	(1)	(2)	(3)	(4)

We may assume the first respondent ticks unimportant and the second respondent ticks important while answering the question. Therefore the ratio of the scale values in the first case is $\frac{4}{2} = 2$ and if we use Scale B, the ratio is $\frac{3}{1} = 3$. Although the differences in the both cases is same and equals two, the ratios are different in both cases and have no meaningful interpretation. The statistical procedure applicable in this case are range, mean, standard deviation and product moment correlation. However, all the statistical procedures applicable for ordinal and nominal scale measurements can also be used in this case.

4. **Ratio Scale :** Ratio scale measurements are the most powerful measurement discussed so far as they pocess all the properties of the measurement scales which we have discussed. Ratios of the numbers on these scales have meaningful interpretation. They possess an unambiguous starting point. The mathematical form of the measurement is written as

$$Y = bX$$

The examples of ratio scale measurement are Income, Distance travelled from home to workplace, Height, Weight, Density etc. The statistical methods applicable for this type of measurement are geometric mean, harmonic mean, coefficient of variation etc. Also all the statistical procedures described for the remaining type of measurements are also applicable under this case.

As mentioned in the preceding paragraphs, the methods of analysis for data depends upon the type of measurements used to measure it. The table below gives the permissible statistics to be used under different types of measurement.

Table 1 : Scales Of Measurement

Scale	Permissible Statistics
Nominal	Mode Contingency coefficient
Ordinal	Median Percentile Rank Order Correlation Sign test Run test
Interval	Mean Average deviation Standard deviation Product-moment correlation t test F test
Ratio	Geometric Mean Harmonic mean Coefficient of variation

Activity 1

What measurement scale is used to determine each of the following?

a) Do you own a washing machine?

_____

b) What is your marital status?

_____

c) How many books are there in your university's library?

_____

d) How satisfied are you with the service of your bank?

_____

e) What is your preference for various brands of toothpaste?

_____

f) Were you present in the last marketing class?

_____

g) How many members of your family would like to watch DD-1 during the prime time?

_____

Activity 2

A survey was conducted to determine the perception about residential carpets in two towns. A sample of 200 respondents was selected from each of the two towns. One of the questions asked to the respondent was the following:

How important is price to you while buying a carpet for your home?

Very important	Important	Neither important nor unimportant	Unimportant	Very unimportant
(1)	(2)	(3)	(4)	(5)

The results of the survey are as follows:

Response category	Percent of respondents checking that category	
	Town 1	Town 2
1	10	05
2	30	20
3	20	10
4	25	40
5	15	25

a) Name the measurement scale being used

.....

b) Find the mode and median for the two samples.

.....

.....

.....

.....

Find the samples means for both the rowns. Can we meaningfully interpret the ratios of two means in this case? If yes, why, if not, why not?

.....

.....

.....

.....

7.4 FORMATS OF RATING SCALES

The rating scales can take a number of physical forms. Suppose the management of a bank A wants to assess the overall customers attitude towards the bank. A number of alternative rating scales can be used by bank. This is exhibited in Table 2 below:

Table - 2 : Different Types Of Rating Scales

1. Please put a ✓ mark at an appropriate place on the line below to express your opinion about Bank A.

   Very Bad _____ Very good

2. Please put a ✓ mark at the category which expresses your overall opinion about Bank A.

   Very bad ... Very good

   1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

3. Please check one of the following category to express your overall opinion about Bank A.

Very bad	Bad		Neither bad nor good		Good		Very good
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]

4. Indicate which of the following summarizes your overall opinion of Bank A?

Terrible	Poor	Fair	Good	Very good	Excellent
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

5. How do you overall rate Bank A in comparison to other banks in your town?

Much worse	Worse	About the Same	Better	Much better
[ ]	[ ]	[ ]	[ ]	[ ]

6. Rank the following by placing a 1 beside the bank you think is best overall, a 2 beside the bank you think is second best, and so on:

Bank A _____ Bank B _
Bank C _ Bank D _____
Bank E_____

7. In each of the following pairs, which bank in your opinion is overall better? (Please check one bank within each pair.)

_____ Bank A or _____ Bank B _____ Bank A or _____ Bank C
_____ Bank A or _____ Bank D ___ Bank A or _____ Bank E
_____ Bank B or _____ Bank C _____ Bank B or _____ Bank D
_____ Bank B or _____ Bank E _____ Bank C or _____ Bank D
_____ Bank C or _____ Bank E _____ Bank D or _____ Bank E

8. Allocate a total of 100 points among the following Banks, depending on how favorable you feel toward each; the more highly you think of each Bank, the more points you should allocate to it. (Please check that the allocated points add to 100.)

Bank A _____ points
Bank B _____ points
Bank C _____ points
Bank D _____ points
Bank E _____ points

100

The type of different formats are given as under:

1) **Graphic & Itemized Formats :** Let us look at question number one and two in the above table. They represent a graphing rating scale which presents a continuum in the form of a straight line. There are infinitely large number of ratings that are possible. The assumption in this rating scale is that it enables the respondent to detect fine shades of differences in attitudes. To measure how favourable the respondent has an attitude towards the bank, one measures the physical distance from the left extreme to the position indicated by him in questions numbering 1 & 2. Higher the distance, more favourable the respondent has attitude towards the bank. Graphic rating scale in question number 1 has ratio-scale properties.

The disadvantages of using a graphing rating scales are as follows:

a) Coding and analysis of data will not only be difficult but also require a substantial amount of time as the researcher firstly has to measure the physical distance on scale for each respondent.

b) Respondents may not be capable of perceiving fine differences in the attitudes on a straight line.

Look at question numbers three, four & five. They represents *Itemized Rating* Scale where there are a set of distinct response categories and the respondent is to choose one among them. The advantages of itemized rating scale are as follows:

a) Coding and analysis of raw data is easier as compared to the graphic rating scale.

b) It is easy for a respondent to answer questions listed under itemized rating scale. Further, they are more meaningful from the point of view of the respondent.

2) **Comparative & Non-comparative Formats:** Let us compare first four questions given in table two with the last four questions. Questions numbering five to eight come under the umbrella of comparative rating scale. They provide the respondent with a common frame of reference for answering the questions. It is assumed that the respondents are answering the same question with a standard frame of reference. However, question numbers one to four represent non-comparative rating scale which permits the respondent to use any or even no frame of reference while answering the questions.

3) **Forced & Non-forced Response Formats:** Compare question numbers three & five with question numbers two, four & seven. In the case of question numbers three and five the respondent is given the choice of being neutral. Such scales are called non-forced choice scales. They have odd number of response categories. The advantage of using this type is that in case the respondent wants to be neutral, one can exercise this option. The problem in this case is that if the respondent has not made up his mind or is not sure about the answer he will tend to tick the middle category thereby preferring to be neutral.

In question numbers two, four & seven; the respondent is forced to take a stand either way. He has no option of remaining neutral. Such types of scales are called forced choice scale. The question which may arise in our mind is that which of the two format is better - the one with odd number or even number of response categories. There is no definite answer. However, the scale representing odd number of response categories are widely used in marketing research because of their advantage of giving respondent the choice of being neutral.

4) **Balanced & Un-balanced Formats:** A balanced scale has an equal number of positive (favourable) & negative (unfavourable) response choices. Questions numbering three and five represent such a scale.

However, if you look at question number four which doesnot have equal number of positive & negative response categories and therefore is biased. This type of scale is called un-balanced scale. The scale in question is biased towards positive (favourable) side. If a respondent did not have favourable opinion about the bank, he is not given many options to express his opinion.

5) **Constant Sum Rating Scale Format:** Let us look at question number eight. In this question the respondent is asked to allocate points depending upon how favourable he feels towards each bank. The minimum points can be 0 and the sum total of point, have got to be 100. Suppose a respondent gives 40 and 10 points to Bank A & B respectively. We can say that for the concerned respondent's opinion of bank A is four times favourable than the opinion of bank B. This type of measurement is called ratio scale measurement which has been discussed before. This scale is widely used in MR studies to know the weightage a respondent gives to various parameters of the service.

6) **Paired Comparison Rating Format:** Let us look at question number seven. There are five banks. We have chosen two banks at a time and the respondent is forced to choose one alternative in each of the ten cases listed there. For example the respondent has to choose between A & B, A & C, A & D and so on. Since there are five banks (items) the number of paired comparison possible are ten. To generalize, if there are n attributes of a product then the maximum number of paired comparison would equal.

$$\left(n \times \frac{(n-1)}{2} \right)$$

Suppose we have a set of five brands (A, B, C, D, E) of a product. Under this method researcher will form all possible pairs of brands. (i.e. ten combinations like AB, AC, AD, AE, etc.) Then each respondent is asked to judge the various combinations of brands on some basis. Finally, the responses are cumulated. Suppose the study shows that 80% consumers say brand A is preferred to brand B.

Further suppose, when brand C was compared with brand A, only 55% of the consumers preferred brand A to brand C. This will tend to imply that the distance between brand A and brand B is much larger than the distance between brand A and brand C. Secondly, if 50% people prefer one brand to another, this will mean that the these two brands are having the same position in terms of the consumers perception.

After cumulating the various responses one may obtain the picture as shown in table 3.

Table 3 : Observed Proportions Saying Brand-I (Column Of Table) is Preferred To Brand (Row of Table)

	Brand - i				
Brand-j	**A**	**B**	**C**	**D**	**E**
A	0.50	0.35	0.25	0.69	0.82
B	0.65	0.50	0.73	0.85	0.93
C	0.75	0.27	0.50	0.16	0.59
D	0.31	0.15	0.84	0.50	0.25
E	0.18	0.07	0.41	0.75	0.50

Note the table shows 65% of the consumers prefer brand A to brand B. That is, the remainder 35% had said brand B is preferred to brand A. Here the diagonal entries have been intentionally filled up with a figure of 0.50 (meaning that there is no difference between these brands).

It can be shown that under assumptions of Normality, these proportions can be interpreted as indicating the distance between those two brands. Using these proportions as probabilities, ordinates from any Normal distribution table for each proportion may be obtained. These ordinates are symbolically represented as Z values. In general, Z values have a symmetric pattern, around Zero. If the proportion is less than 0.5, the corresponding Z value has a negative sign, and conversely, if the proportion is greater than 0.5, the Z value is positive. The Z values for the above mentioned data are as follows.

Table 4 : Z Values Corresponding Top Earlier Preference Data

	Brand - i				
Brand-j	**A**	**B**	**C**	**D**	**E**
A	0	-0.39	-0.67	0.50	0.92
B	0.39	0	0.61	1.04	1.48
C	0.67	-0.61	0	-0.99	0.23
D	-0.50	-1.04	0.99	0	-0.67
E	-0.92	-1.48	-0.23	0.67	0
Total	-0.36	-3.52	0.7	1.22	1.96
Mean (Z)	-0.072	-0.704	0.14	0.244	0.392
R	0.632	0	0.844	0.948	1.096
Brand	B	A	C	D	E
Interval Scale Value	0	0.63	0.84	0.95	1.096

In this manner one can obtain an interval scale for the positions of different brands from the original data which were ordinal in nature.

7) **Rank Order Scale Format:** Let us look at question 6. The data generated by this question has ordinal scale measurement. The rank order technique is comparative in nature in the sense that respondents are asked to rank comparatively the items listed.

The advantage of using rank-order scale lies in its easiness to tabulate the results. Moreover, it is easy and less time consuming to administer than paired comparison test which is also a comparative rating technique. Further, it forces the respondents to discriminate among products in realistic manner.

This technique has several limitations. The data obtained could be misleading if all the alternatives which form respondent preference are listed out. Also, if the respondent "dislikes" all the attributes listed, the one ranked first is the least "disliked" attribute. The researcher therefore must ensure that a realistic set of objects are being evaluated and should be sensitive to the attitudinal position of the respondent. Finally, ranked-order scale do not reveal how far apart the attributes stand in the perception of respondents and why were they ranked that way.

Activity 3

Construct a constant sum rating scale to indicate the factors which prespective management students choose while selecting a management institute for pursuing a management programme.

.....

.....

.....

.....

.....

.....

.....

.....

Activity 4

Why is that balanced response format preferred to unbalanced response format?

.....

.....

.....

.....

.....

Activity 5

Take any four brands of toilet soap and administer a paired comparison rating scale on a set of 100 respondents. Show how would you obtain an interval scale for the positions of different brands from the original data.

.....

.....

.....

.....

.....

.....

7.5 SINGLE ITEM VS. MULTIPLE ITEMS SCALES

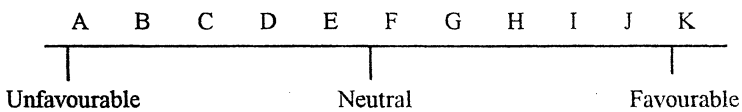
Single Item Scale : A single item measures the opinions of the respondents through just one rating scale. For example, questions numbers 1 -5 in table 2 are single item rating scale.

Multiple Item Scale : In the multiple item scale a number of statements are used pertaining to attitude towards an object. Each statement contains a rating scale attached to it. The disadvantage of single item rating scale is that it is a crude measure of the feelings/opinions of the respondent, whereas, this is not the case with multiple item scales. Below we mention the following widely used multiple item scale.

1. Thurstone Equal Appearing Interval Scale
2. Likert Scale
3. Semmantic Differential Scale (Bi-polar Scale)
4. Fishbein's Scale

1) **Thurstone Equal Appearing Interval Scale :** This scale of attitude measurement was first described by Thurstone and Chave. To construct the scale one needs to follow the steps outlined below:

- A large number of statements pertaining to the subject of enquiry are collected. This is done through exploratory research by using existing literature on the subject, discussion with knowledgeable persons, personal experience & through focus-group interviews.
- The statements range from one extreme of favourable attitude to other extreme of unfavourable attitude.
- The number of statements should be large.
- Each statement is written on a separate card and subjects are asked to sort these statements into a number of intervals.
- The number of intervals could be eleven put on cards labelled A to K where A corresponds to most unfavourable attitude, K corresponds to most favourable attitude and F represents neutral attitude.



- The piles bearing letter from A to K could be regarded as having numbers from 1 to 11 indicating that the respondent has rated each statement on a 11-point rating scale. Therefore 11-point rating scale becomes the psychological continuum on which statements have been judged.
- An average value is calculated in respect of each statement. This value is treated as the scale value (median value for the example to follow) of the statement. Let us consider the data given in the table 5.

Table 5 : Summary Table For Judgements Obtained By The Method Of Equal - Appearing Intervals

		Sorting Categories												
Statement		A 1	B 2	C 3	D 4	E 5	F 6	G 7	H 8	I 9	J 10	K 11	Scale Value	Q Value
1	f	2	2	6	2	6	62	64	26	18	8	4	6.8	1.7
	p	.01	.01	.03	.01	.03	.31	.32	.13	.09	.04	.02		
	cp	.01	.02	.05	.06	.09	.40	.72	.85	.94	.98	1.00		
2	f	0	0	0	10	40	28	50	26	28	14	4	6.9	2.8
	p	.00	.00	.00	.05	.20	.14	.25	.13	.14	.07	.02		
	cp	.00	.00	.00	.05	.25	.39	.64	.77	.91	.98	1.00		
3	f	0	0	0	2	8	6	26	44	56	44	14	8.7	2.0
	p	.00	.00	.00	.01	.04	.03	.13	.22	.28	.22	.07		
	cp	.00	.00	.00	.01	.05	.08	.21	.43	.71	.93	1.00		

Source: Edwards, Allen L., Bombay, Vaklis, Feffer and Simons Pvt. Ltd., 1969 (Indian Reprint), p.87.

- The table 5 gives data in respect of three statements administered for 200 respondents. Corresponding to each statement there are three rows; the first row indicates the number of respondents placing the statement in respective categories (f), the second row shows the probability of occurrence computed as the ratio of frequency for that category to the total number of respondents (p), the third row indicates the cumulative frequencies given by (cp).

Scale value or Median value $$S = l + \left(\frac{.50 - \Sigma P_b}{P_w}\right) \times i$$

l = lower limit of the interval in which the median lies

ΣP_b = the sum of proportion below the interval in which the median lies

P_w = the proportion within the interval in which the median lies.

i = the width of the interval

The above formula can be applied to statement number 1 to get the scale value (S) as follows:

$$S = 6.5 + \left(\frac{.50 - .40}{.30}\right) 1.0 = 6.8$$

Similarly the scale value corresponding to statement number 2 and 3 can be computed as 6.9 and 8.7 respectively. To find the interquartile range, one needs to compute the first and third quartile of the distribution. The interquartile range denoted by $Q = Q_3 - Q_1$ where Q_3 = third quartile and Q = first quartile. The third and the first quartile of the distribution corresponding to statement number 1 is given below:

$$Q_3 = 1 + \left(\frac{.75 - \Sigma P_b}{P_w}\right) i$$

$$= 7.5 + \left(\frac{.75 - .72}{.13}\right) 1.0 = 7.7$$

$$Q_1 = 1 + \left(\frac{.25 - \Sigma P_b}{P_w}\right) i$$

$$= 5.5 + \left(\frac{.25 - .09}{.31}\right) 1.0 = 6.0$$

Therefore interquartile range Q is given by

$$Q_3 - Q_1 = 7.7 - 6.0 = 1.7$$

- A small value of Q indicates that there is agreement among the respondents in judging the degree of favourableness or unfavourableness of a statement. Similarly, using the same argument we can infer that a large value of Q indicates disagreement among the respondents in judging the degree favourableness or unfavourableness of a statement. Thurstone & Chave treated the statements with large value of Q as ambiguous.
- This way about 20 to 30 statements are selected. When two or more statements have the same Q value and only one is to be retained, the one having the lowest value of scale is selected.
- After having selected 20 to 30 statements, they are embodied in the form of a questionnaire to be administered to the respondents. The statements are arranged in a random sequence. Each respondent is asked to select those statements which best reflect his or her feelings towards the attitude object. The average of the score values of all statements which the respondent endorses becomes his scale score. Assuming that the respondent endorses statement numbers 1, 2 & 3 (Table 5), his score will be

$$\frac{6.8 + 6.9 + 8.7}{3} = \frac{22.4}{3} = 7.5$$

As we know that the neutral attitude has a score value of 6 on a 11-point scale, a score of 7.5 indicates that the respondent has a favourable attitude towards the object.

Limitations of Thurstone Scale: Thurstone scale uses a two-stage procedure and therefore it is both time consuming and expensive to construct. As there is no explicit response to each item, it does not have much diagnostic value. The scale has also been criticized on account of a method of scoring. Respondents are merely asked to select those statements with which they agree. Therefore, there is a possibility of two or more respondents having the same attitude score. For example, if the respondent A agrees with statements having a score value of 5, 7, 10 and another respondent B agrees with statements having scale values of 8, 8 & 6, both of them will have the same attitude towards the object which may infact not be true.

2) **Likert Scale:** This scale comprises of a series of evaluative statements/items concerning an attitude object. Each statement has a five point agree disagree scale. The number of statements/items could vary from study to study. However a typical likert scale has generally 25 to 30 items. The scores on the individual items are summed to produce a total score for the respondent. For this reason it is also called a summated scale. The assumption made here is that each of the items/ statements measures some aspect of a single common factor otherwise the items cannot be legitimately summed.

Let us recall the problem we discussed while talking about formats of different types of scale. Assuming the management of the Bank A wants to know the opinion about the bank from its customer. The table 6 below gives a set of six statements from a likert scale that may be used to measure customers attitude towards bank.

Table 6 : Likert Scale Items

	Strongly disagree	Dis-agree	Neither agree or disagree	Agree	Strongly agree
1. Counter clerks at the bank are friendly.	—	—	—	—	—
2. Lines at the counter move slowly.	—	—	—	—	—
3. The bank keeps confi-dentiality of the transactions.	—	—	—	—	—
4. The bank offers a variety of products.	—	—	—	—	—
5. The bank's operating hours are inconvenient.	—	—	—	—	—
6. The ATM facilities of the bank break down frequently.	—	—	—	—	—

To construct a likert scale we proceed as follows:

1) Generate a number of statements relevant to the attitude. They should be either clearly favourable or unfavourable. There are no rules for generating initial set of statements. It may be done through exploratory research by talking to knowledgeable people and having focus group interviews. Knowledgeable people may include the employees and the customers of the company. Managerial judgement could help in generating a broad range of items. A lot of care needs to be taken in choosing statements. In order to capture all relevant aspects of attitude object, it is essential to include sufficiently varied statements.

2) Once items are written, they are administered on a sample which is representative of the population being studied.

3) It may be noted that the response categories have verbal and no-numerical labels. This is usual format adopted in which Likert scale statements are presented to respondents. However, after the scale is administered, the numbers are assigned to responses in order to generate a quantified measure of attitudes. The numbers one through five are normally used for this purpose. Other sets of numbers like +2, +1, 0, -1, -2, can also be employed.

   By examining the items in table 6, one may note that statement numbers 1, 3, and 4 are favourable towards the bank while the remaining are unfavourable. It is essential that a good Likert scale must have equal number of favourable and unfavourable statements. The "strongly agree" category attached to favourable statements and the "strongly disagree" category attached to the unfavourable statements must both be assigned the same number (e.g., the number 5, if a 1-through 5 numbering scheme is used).

As discussed before, to design a good Likert scale one must first of all generate a large pool of statements relevant to the measurement of an attitude. Then statements that are vague and/or non-discriminating should be eliminated.

Suppose management of bank wants to develop a 20-item scale to measure customer attitude towards bank. To start with, a large number of items, say 100, similar to the six statements shown in table 6 are developed. The following procedure may be used in reducing the set of initial 100-items to a final 20-item instrument.

A questionnaire with initial 100-items with 5 possible response categories ranging from strongly disagree to strongly agree is administered on a set of selected representative respondents. The respondents are asked to rate the bank on each statement by ticking one of the five response categories. The categories are numbered from 1 to 5 with higher number showing the more favourable attitude towards the bank. As there are 100-items the minimum and maximum score for a respondent would vary from 100 to 500. This is possible as one can sum up the scores of individual items to produce a total score for the respondent under the assumption that each item/statement measures

some aspect of a single common factor. Therefore higher the score for a respondent, the more favourable attitude he has towards the bank in question. This shows the total score plays a very important role in reducing the initial set of 100-items to an appropriate scale comprising of 25 to 30 items.

To illustrate, consider the following total score of two respondents, A and B, as well as their scores on two individual items, i and j.

Respondent	Score on item i	Score on item j	Total Score
A	3	5	480
B	3	1	170

From the above table we may conclude that respondent A has a more favourable attitude towards the object than respondent B. Further by examining at their scores on item i and j we find that item i is a poor discriminator while item j is a good discriminator between two respondents. In other words if we compute the correlation coefficient between the scores on item i with total score and score on item j with total score, we will find that the later will have a high degree of correlation. This logic may be extended across all the statements. Therefore those items on which respondent scores correlate highly with the respondents total score are better for inclusion in the final scale than other items.

It is important to understand the difference between Likert scale & Thurstone scale. In Thurstone scale a respondent is expected to endorse those statements which best reflect his or her feelings towards the attitude object. However, in Likert scale a respondent is to answer every item/statement and a high total score indicates a high favourable attitude towards the object.

Likert scale is less time-consuming and less laborius than the Thurstone scale. It is seen that the reliability coefficient (to be discussed in 7.6) is more for Likert scale than for Thurstone scale.

3) **Semantic Differential Scale:** This is also called bi-polar scale. The scale may be used in cases such as comparison of brands, comparison of companies images and to determine the effectiveness of advertising on attitude change etc. The scale is similar to Likert scales as it consists of a series of items to be rated by respondents. However, there are differences between Likert scale and Semantic differential scale.

The Semantic differential scale comprises of a series of bi-polar adjectives or phrases that pertain to the attitude towards the object. (Likert scale had complete statements.) Each pair of opposite adjectives is separated by a seven category scale (at times with five category or nine category scale). No numerical labels or verbal labels other than anchor labels are used for representing various categories. Some of individual scales have favourable descriptors on the right-hand side while the others have it on the left-hand side. The rational for having this sort of an arrangement is similar to that of Likert scale.

Respondents are asked to put a cross on one of the seven category that best describes his/her views about the attitude object along the continuum implied by bi-polar adjectives. Going back to the example of bank where the management is interested in finding the customer's opinion about bank A, adjectives representing the same dimensions as described by Likert scale in table 6 is given below.

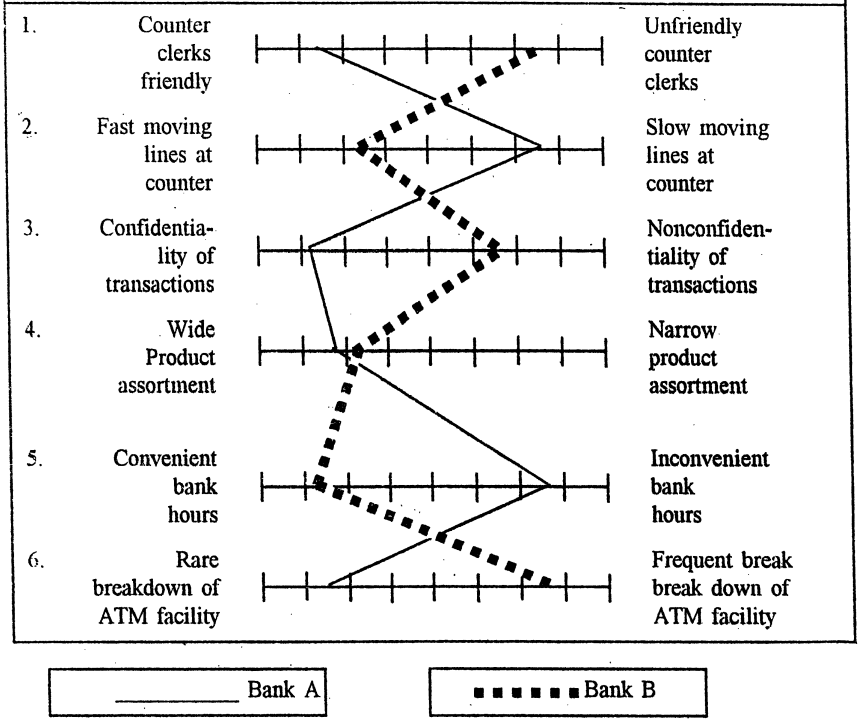
Table 7 : Semantic Differential Scale Items

1.	Counter clerks friendly	Unfriendly counter clerks
2.	Slow moving lines at counter	Fast moving lines at counter
3.	Confidentiality of transactions	Nonconfidentiality of transactions
4.	Wide product assortment	Narrow product assortment
5.	Inconvenient bank hours	Convenient bank hours
6.	Frequent break down of ATM facility	Rare breakdown of ATM facility

The scale is administered to a set of respondents and corresponding to each respondent his total score is computed. This is possible as the categories can be numerically coded from 1 to 7, 7 being the highest value given to the favourable adjective. One should make sure to reverse the code items that are reversed as was done in Likert scale. Also the overall attitude score can be computed by summing the coded responses on the individual items and this also has the same interpretation as was the case of Likert scale.

Semantic differential scale helps us to develop a pictorial profile of the attitude objects based on the mean ratings on the individual items. This is shown below in table 8 for two banks namely bank A and bank B. To facilitate comparison, it is customary to place all favourable discriptors on the same side of the diagram.

Table 8 : Pictorial Profile Based On Semantic Differential Ratings



We note from above table that Bank A has a significant edge over Bank B with respect to friendly nature of counter clerks, confidentiality of transcations and smooth functions of ATM facility. Bank B is, however, preceived better than bank A in terms of convenient working hours and speed at the counter. Bank A is perceived to be better than B in terms of product assortment; however advantage to Bank A does not seem to be significant. These results can pinpoint a particular bank's relative strength and weakness as perceived by customers.

4) **Fishbein's Scale :** This scale uses a combination of constant sum and semantic differential scale. To assess how consumers **perceive a set of** alternative offerings, this type of scale is used. Here the respondent **is first asked to assign a weightage** to a set of given attributes. Afterwards the **respondent is given some specific** choice alternatives. He/she is asked to judge **these alternatives on those specified** list of choice criteria. Ultimately one can compute a composite score for each alternative. For example, suppose a research study **was** assessing consumers' attitude toward five leading brands of decorative **paints in terms** of the attributes like Durability, Washability, Finish, Range of **package size** Range of colour, Price, Ease of availability etc.

Here the respondent is first of all asked to evaluate the important of the given attribute in the paint choice decision.

Secondly, he is asked how strongly do you rate the brand X on the given attribute

Ultimetely each consumer's attitudinal score for brand X (AX) is found by the following formula

$$Ax = \sum_{i} Wi\ R_i\ x$$

Where Wi = Consumer's ratings of the ith attribute.

Rix = Consumer's judgement of brand X on the i^{th} attribute.

Similarly, one can compute the consumer's **attitudinal score for other brands of** paint examined. Ultimately, by pooling the **scores of different gruops of consumers** one can assess the consumer's perception **about various paints**

Fishbein scale is found useful in problems **such as comparing brand image and** advertising designs. In general, it provides **a convenient way of assessing consumer** attitudes.

Activity 6

Construct a Scmemantic differential scale to measure the perception of a brand of office from a set of consumers.

.....

.....

.....

.....

.....

.....

Activity

The choice of a PC is taken on the evaluation of different considerations. Prepare a list of features that could be rated on Likert Scale.

.....

.....

.....

.....

.....

.....

Activity 8

List out the merits of Likert scale over Thurstone equal appearing interval scale.

.....

.....

.....

.....

.....

7.6 CRITERIA FOR GOOD MEASUREMENT

The adequacy of scale is judged in a way as to have measurement of acceptable quality. The overall usefulness of the scale depends upon its validity, reliability and sensitivity. These are discussed below:

1) **Validity** : A scale is said to be valid if it measures correctly what it is expected to measure. As we know attitudes are multifaceted, therefore single item scales are usually deficient on this criterion. In other words an attitude scale is valid only when it is real and correct. There are several different types of validity as discussed below.

   a) **Content Validity** : This is also known as face validity. Here the contents of an attitude scale should cover all relevant facets of an issue which influence respondent's attitude. Suppose we have a scale to measure job satisfaction of employees in an organization. The scale covers various dimensions like nature of work, pay, security and superior. The scale, however, lacks in content validity as it does not cover an important dimension like company policies and practices. This could happen because what dimensions to be included in a scale would depend on the judgement of the researcher which is likely to vary from individual to individual. Therefore, in order to avoid this, it would be better to approach a group of knowledgeable persons rather than leaving to one person.

   b) **Construct Validity** : It is known that it is not possible to measure attitude directly. It is inferred indirectly from the responses given by the respondents. Construct validity involves understanding the theoretical rationale underlying the obtained measurements. The content validity of an attitude can be assessed quantitatively by finding its correlation with measures of other constructs that one would expect to be strongly associated with the attitude and measures of constructs that would not be correlated to the attitude. Strong correlation of attitude with the construct is an example of convergent validity whereas low correlation of attitude with the construct is an example of discriminating validity. A scale with construct validity should have both convergent and discriminating validity.

c) **Predictive Validity :** It involves the ability of a measured market phenomenon at one point of time to predict successfully another market phenomenon at another point of time. If the correlation between two is high, the initial measure is said to have a high predictive validity. An opinion questionnaire which forms the basis for correctly forecasting, the demand for product has high predictive validity.

2) **Reliability :** A scale is said to be reliable when it gives the same measurement under similar condition. If a scale makes equal error every time, it would be reliable. However such a scale can not be valid as we know that the validity depends upon correct measurement. Reliability is achieved when the scale is free from erratic measurements. The following two methods are used for testing reliability.

a) **Test-Retest Reliability :** It is concerned with how stable the ratings are when the scale is administered to the same group of persons at two different points of time. If there is a high correlation between two sets of scores (consistency), the test - retest reliability is very high.

One should be cautious while using this method to test reliability. If the time difference between two sets of observations is long, it is likely that the attitude may undergo a change. Further if the time difference is too small, the respondents are likely to remember their earlier responses and therefore memory effect may distort the reliability test. There are no guidelines in determining what should be the ideal time interval between two observations so as to take care of these problems. A rule of thumb is to use the time interval between two to four weeks.

b) **Split-Half Reliability :** It can be assessed only for multiple-item scales. Here the scale items are split randomly into two equal parts. At times splitting may be done by putting even number of items on one side and odd number on the other side.

The correlation coefficient between respondents' total score derived from two sets of items is computed. A high degree of correlation indicates a high split-half reliability of the scale.

3) **Sensitivity :** The sensitivity of a scale is closely related to its reliability. It is the ability of a scale to be able to discriminate between respondents who differ even slightly in terms of their attitude toward something. An essential of a rating scale is that it should have a sufficient range of numbers to help detection of fine variations in attitude.

Activity 9

Give an example to show that a scale could be reliable yet it may lack validity.

.....

.....

.....

.....

.....

.....

Activity 10

List out the limitations of test-retest reliability.

.....

.....

.....

.....

.....

.....

Activity 11

What precautions a research should take to ensure that the scale developed by him/her does not lack in content validity?

.....

.....

.....

.....

.....

.....

7.7 SUMMARY

Measurement involves the use of numbers to represent the marketing phenomenon under investigation. The direct measurement of attitude is a very difficult task. It can only be inferred indirectly from the responses of the respondent. Attitude comprises of three components viz a cognitive component, an affective component and a behavioural component. For marketing decisions the attitude behaviour relationship relates to measuring of cognitive and affective components so as to be able to predict future behaviour.

There are four types of scales used in marketing research to infer attitude towards a particular product /service. There are Nominal, Ordinal, Interval and Ratio scale. The most powerful among them is ratio scale measurement. A study of these scales is important in the sense that the method of analysis for data depends upon the type of measurement used to measure it.

The rating scale may take various physical forms. They could be presented in graphic and itemized formats, comparative and non-comparative formats, forced and non-forced response formats, balanced and un-balanced formats, constant sum rating format, paired comparison rating format and ordinal scale format.

Further the attitude scale could be single item or multiple item scale. A single item measures the opinion of the respondents through just one rating scale whereas for multiple items a number of statements pertaining to attitude towards objects are used. The limitation of the single item scale is that it is a crude measure of the feelings/ opinion of the respondent. A number of widely used multiple item scales namely Thurstone Equal Appearing Interval Scale, Likert Scale, Semmantic Differential Scale and Fishbein's Scales with their limitations and uses are discussed.

The unit is concluded by discussing the criteria of a good measurement scale. Three criterion namely validity, reliability and sensitivity are used for the purpose.

7.8 SELF-ASSESSMENT QUESTIONS

1) The centigrate fahrenheit relationship is given by

$$C^{o} = \frac{5}{9}(F^{o} - 32)$$

show that in this scale the ratio of two scale values have no meaningful interpretation.

2) Suppose management of an insurance company wants to ascertain the perception of its customers towards the company. Construct a 7-item scale to measure the perceived perception of the company using Likert Scale and Semmantic Differential Scale. Make sure that the 7-items under each of the two scales correspond to the same seven dimensions.

3) What are the four different levels of measurement? Discuss the mathematical operations which may or may not be used under each level of measurement.

4) Explain the three criterion of measuring the usefulness of an attitude scale.

5) Distinguish between graphic and itemised rating scale. What are the advantages of itemised scale over graphic rating scales? Name the level of measurement which corresponds to the data collected by graphic rating scale.

6) Distinguish between the following:
   a) Forced & Non-forced response formats
   b) Comparitive & Non-comparitve formats
   c) Balanced & Un-balanced formats

7) Show with the help of an example how you can convert ordinal scale measurement to nominal scale measurement.

8) The method of analysis of data depends upon the level of its measurement" Discuss.

9) Indicate giving reasons the scale of measurements for the following variables:
   i) Geographical area
   ii) Costs
   iii) Index numbers
   iv) Preferences
   v) Cast in a society.

10) Distinguish between the validity and reliability of a measure. How may the reliability of a measure be evaluated?

11) Explain how ordinal and interval scale can be derived from paired comparison data. What advantages and limitations exist for the paired-comparison technique?

7.9 FURTHER READINGS

Beri, G. C. "Marketing Research - Text and Cases" Tata McGraw-Hill Publishing Co. Ltd.

Bhatnagar, O.P. "Research Methods and Measurements in Behavioural and Social Sciences" Agricole Publishing Academy.

Green, Paul E. and Donald S. Tull "Research for Marketing Decisions" Prentice-Hall of India Pvt. Ltd.

Guilford, J.P. "Psychometric Methods" Tata McGraw-Hill Publishing Co. Ltd.

Kinnear, Thomas C. and James R. Taylog, "Marketing Research - An Applied Approach" McGraw-Hill International Editions.

Luck, David J. and Ronald S. Rubin, "Marketing Research" Prentice-Hall of India Pvt. Ltd.

Majumdar, Ramanuj "Marketing Research - Text, Applications and Case Studies" Wiley Eastern Ltd.

Parasuraman, A. "Marketing Research" Addison-Wesley Publishing Co.

Westfall, Boyd and Stasch. "Marketing Research - Text and Cases" Richard D. Irwin. Inc.

NOTES

NOTES

NOTES

Uttar Pradesh
Rajarshi Tandon Open University

MBA

Marketing Research

Block

3

DATA PROCESSING AND ANALYSIS

UNIT 8
Qualitative Research 5

UNIT 9
Data Processing — Coding, Tabulation And Data Presentation 16

UNIT 10
Description and Inference from Sample Data 36

UNIT 11
Analysis of Association 58

Expert-Cum-Course Writers Committee

Prof. Rakesh Khurana, *Chairman*
School of Management Studies
IGNOU, New Delhi

Prof. Abraham Koshy
Indian Institute of Management
Ahmedabad

Prof. Y.L. Grover
Management Consultant
New Delhi

Prof. Deepak Chawla
International Management Institute
New Delhi

Prof. Mukul Gupta
International Research Bureau
New Delhi

Mrs.Neerja Wable
Indian Market Research Bureu
New Delhi

Shri Saurabh Khoshla
MD, Tulika Advertising Agency
New Delhi

Prof. Sudas Roy
Indian Institute of Management
Calcutta

Dr. N Kinra
Indian Institute of Management
Lucknow

Prof. A.P. Arora
Management Development Institute
Gurgaon

Dr. Sanjay K. Jain
Delhi School of Economics
University of Delhi, Delhi

Mrs. Monica Khanna,
New Delhi

Dr. P.S. Vishwanathan
Hindustan Computer Ltd.,
New Delhi

Prof. Sharad Sarin
XLRI
Jamshedpur

Prof. Mohan Aggarwal
XLRI
Jamshedpur

Mr. D.K. Bose
Media Director
R.K. Swamy, BBDO, New Delhi

Dr.(Mrs.) Madhulika Kaushik
Director
School of Management Studies
IGNOU, New Delhi

Prof. R.K. Grover
School of Management Studies
IGNOU, New Delhi

Mr. Qasemuddin Haider
Asst. Regional Director
IGNOU, Patna

Mr. Ravi Shanker
School of Management Studies
IGNOU, New Delhi

Dr. Kamal Yadava
Course Coordinator
School of Management Studies
IGNOU, New Delhi

Print Production: B Natarajan, Copy Editor, SOMS, IGNOU

September, 2000 (Reprint)

Indira Gandhi National Open University, 1998
© ISBN-81-7605-355-4
All rights reserved. No part of this work may reproduced in any form, by mimeograph or any other means, without permission in writing from the Indira Gandhi National Open University. Further information of the Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi - 110 068.

Reproduced and reprinted with the permission of Indira Gandhi National Open University
by G. S. Shukla, Registrar, U.P.R.T.Open University, Allahabad (December 2017)

Printed by : Chandrakala Universal Pvt. Ltd. 42/7 Jawahar Lal Nehru Road Allahabad

211002

UGSEC-02/122

BLOCK 3 DATA PROCESSING AND ANALYSIS

This block consists of four units i.e. Unit 8 to Unit 11. Unit 8 on `Qualitative Research gives you an understanding of how qualitative research differs from quantitative research, its applications in the area of marketing and various methods of conducting qualitative research.

Unit 9 explain the process of coding, tabulation and data presentation. Unit 10 titled `Description and Inference from Sample Data' helps you in distinguishing between descriptive and inferential statistic, explains the usefulness of universal tables and also discusses the various inferential techniques applicable to univariate analysis. The last unit of the block gives the details of `Analysis of Association'.

UNIT 8 QUALITATIVE RESEARCH

Objectives

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

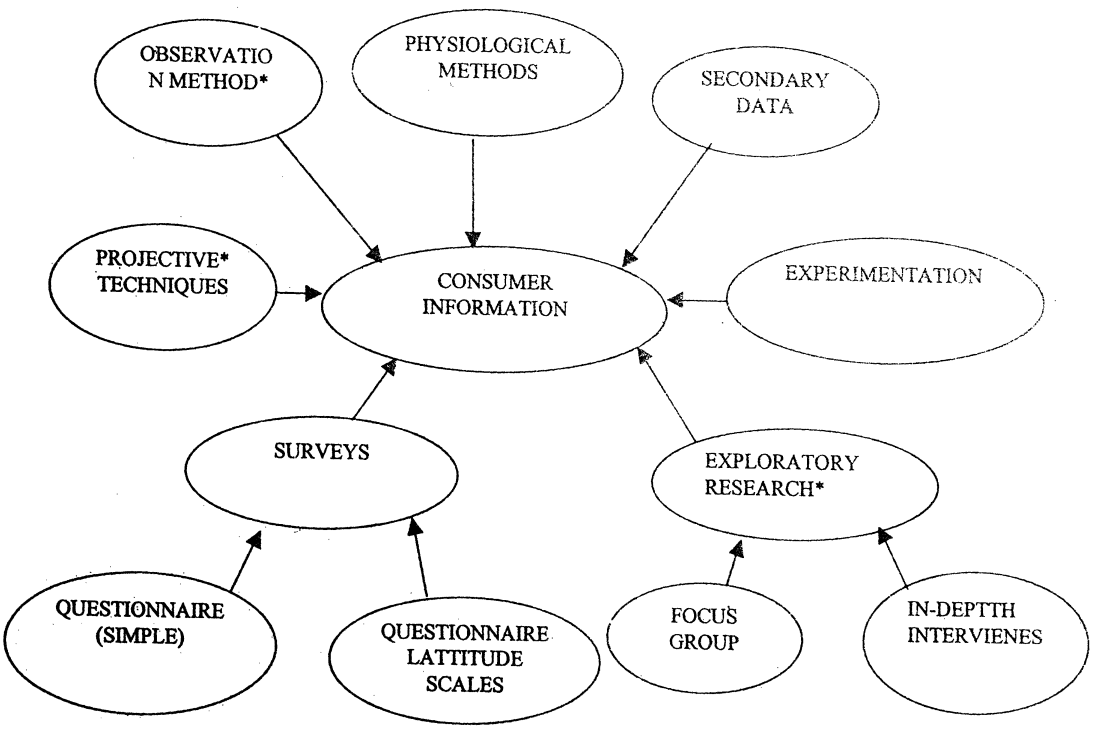
- Differentiate between qualitative research and quantitative research
- Identify possible applications of qualitative research in the area of marketing
- Understand various methods of conducting qualitative research.

Structure

8.1 Introduction to Qualitative Research
8.2 What is Qualitative Research?
8.3 Differences between 'Qualitative' and 'Quantitative' Research
8.4 Uses of Qualitative Research
8.5 Methods of Conducting Qualitative Research
8.6 Latest Techniques of Qualitative Research
8.7 Summary
8.8 Self -Assessment Questions
8.9 Further Readings

8.1 INTRODUCTION TO QUALITATIVE RESEARCH

There are various methods by which information about consumer can be obtained, and they are as follows:

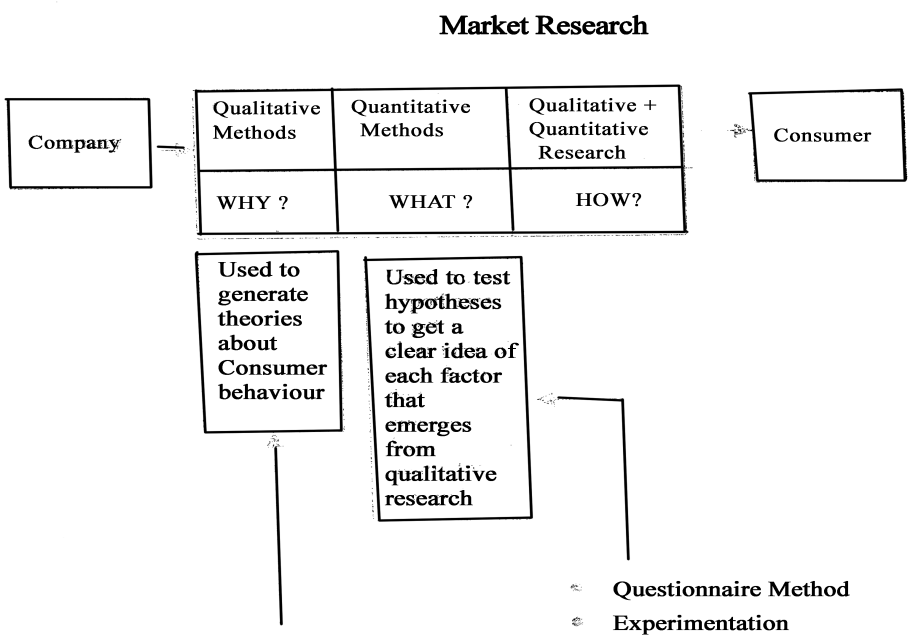


Some of the methods listed out fall in the category of quantitative research and some under the method of qualitative research..(* Marked)

8.2 WHAT IS QUALITATIVE RESEARCH

1) Qualitative Research is exploratory or diagnostic in nature.
2) It involves a small number of people who are not sampled on any probabilistic basis.
3) Qualitative Research is impressionistic rather than definitive.
4) Qualitative Research is used to generate hypothesis for further research.
5) Qualitative Research is used to get better insights into consumer behaviour, and to understand underlying behaviour of the consumer in the buying process .
6) Through Qualitative Research,one can get subtle clues about products/brands/services that very few quantitative studies can replicate
7) In Qualitative Research, no attempt is made to draw hard and fast conclusions about facts that emerge
8) In a market research project, one nomally uses qualitative research and quantitative research methods together.
9) Though Qualitative Research is useful to saidy consumer behaviour any attempt to generalize the finding for the entire population is very risky as the findings are based on a very small sample size
10) Qualitative Research is highly subjective in nature and one would need trained psychologists and interviewers to conduct the research work and also for analysis and interpretation of data

The diagram below explains the importance of Qualitative and Quantitative Research methods.



8.3 DIFFERENCE BETWEEN QUALITATIVE RESEARCH AND QUANTITATIVE RESEARCH

Qualitative Research	Quantitative Research
Exploratory or diagnostic in nature - used to understand behaviour and generate hypothessis	Once hypothesis has been generated, used to test out hypothesis
No calculation of sample size possible - size of sample depends upon time available to conduct research, cost, variation in the population.	Generally a probabilistic approach is used to calculate sample size using the sample size formula.
Sample selected is such that it represents different sections of the population	Random selection of respondents to be part of research work-may or may not represent different population segments, depending upon the sampling method utilized.
It is dangerous to generalize conclusions for the entire population.	Conclusions are generalized to the universe, of which the sample purports to be representative.

8.4 USES OF QUALITATIVE RESEARCH

1. It is used in 'concept generation and evaluation' e.g., whenever company wants to develop new product, or modify existing product, it would conduct Qualitative Research among target consumers to understand new emerging consumer needs, or problems which consumer has with the existing product. Therefore, every time when one goes into the market place, one would find newer models of the same product with addition/deletion of features, new and better external appearance, better finish of the product etc.
2. It is used to define the problem areas more fully - in marketing research one normally starts with qualitative research which is validated further by quantitative research.
3. It is used to formulate hypothesis for further investigation/quantification.
4. It is used to obtain large amount of data about beliefs, attitudes, etc. as data input for developing questionnaires, attitude scales, which would be used as input for multi-variate analysis studies.
5. It is also used to conduct post-research study i.e., to amplify or explain some points that emerge from a major study, without having to repeat on a large scale.
6. In some areas of marketing research — especially understanding of consumer behaviour, a more flexible approach is required, rather than a rigid approach as provided by a structured questionnaire — hence qualitative research could be used.
7. In studies of distribution channels, sales, pricing strategies — quantitative approach is most suitable, whereas in concept development, product development (needs of a consumer), advertising research — qualitative approach is more suitable.

Ultimately, whether one uses a Qualitative Research or Quantitative Research approach, depends upon the objective of the research work, and the value of decisions that would be taken from the data generated by the research work.

8.5 METHODS OF CONDUCTING QUALITATIVE RESEARCH

1. Individual 'Depth' or 'Intensive' Interviews

The in-depth interviews could be classified as:

- Non-directive interview
- Semi-structured or focussed interview.

In a non-directive interview, the respondent is given maximum freedom to respond in a manner that he wishes to, within a reasonable limit of relevancy to the topic under discussion. However, the interviewer retains the initiative in the interview process, else the focus of the interview would be lost. Thus, with this technique, the respondent is given a chance to freely express his ideas and thoughts, which acts as an important feedback to the company regarding the company's products/service.

In a semi-structured or focussed interview, the initiative is retained by the interviewer, and the interview has to cover specific list of points, which has been decided in advance. There is also a tighter control over the interview, in order to maximise data collection and also collect data relevant to the topic under consideration. The interviewer also has determined as to which questions are to be asked. The best example to highlight this interview process is the chat shows that take place on television. Even though the participant(s) is/are given maximum freedom with respect to his/her answers, the initiative is retained by the interviewer, and he/she has decided in advance the questions that would be asked in the course of the interview.

In-Depth interviews are appropriate in the following situations:

1) When detailed probing of an individual's behaviour, attitude, and needs is required.

2) When the subject matter is of a highly confidential nature (e.g., how do you plan your investments required for annual tax planning).

3) When the subject matter is an emotionally charged one or of an embarrassing nature (e.g., how do you spend when you go on a date).

4) When a step-by-step understanding is required of complicated decision making (e.g., how does a family plan its holiday — selection of vacation site, mode of travel and stay places, or how does a family decide when purchasing a house — which is normally a lifetime decision).

5) When interviews are conducted with highly qualified professionals (e.g., surgeous on the usage/problems with various medical equipments), a normal questionnaire would not suffice for getting information, and a detailed probing is required which would come out only through an in-depth interview.

6) When a respondent may not be able to express his/her true feelings in a group discussion. Very often in a group discussion, a few aggressive people tend to dominate the discussion, hardly giving an opportunity to others in the group to express their views. To avoid such a situation, a researcher may decide to conduct an in-depth interview instead of a group discussion. Also, when the topic of the group discussion is of a personal nature and socially embarrassing, people in the group would tend to conform to the views of the overall group, rather than express their own individual ideas.

Limitations of the in-depth interview:

1) The skill of the interviewer is very critical in drawing out the respondent's true feelings.

2) In an interview, there is not only verbal communication, but also non-verbal communication, and the interviewer should also keep track of the respondent's voice tone, facial expression, movement of hands (gestures).

3) The sample size cannot be large, as conducting in-depth interviews takes a longer time.

4) Analysis and interpretation of data is a highly subjective process, thus generalizations to the entire population cannot be done.

5) In-depth interviews act as a source of generation of hypothesis, development of questionnaire for further research work.

Activity 1

Conduct a depth interview withe one person to find out his/her feelings on the purchase and consumption of (a) Packaged fruit juices and (b) Ice- creames.

.....

.....

.....

.....

2. Focus-Group Discussion

There are broadly two ways in which a group discussion can be conducted:

i) **Brain-Storming :** In such a method, there is no moderator for the group, and the group freely expresses its ideas on the given topic. The ideas could be absolutely abstract, but then this would help in generating new product ideas and also better ways of conducting a particular business etc. In this, use is made of tape-recorder to record the group discussion, video-taping of proceedings is also done in order to record the facial expressions of the participants, as also the intensity of their feelings.

ii) **Focussed Group Discussion :** In such a method, the group is given a topic and asked to discuss the topic. A moderator would also be involved in order to ensure that the group discussion remains relevant and does not go off the track. The moderator could stop the discussion between time intervals to find out what conclusions are being drawn by the group after each time interval.

Size and Composition of a Group Discussion Panel :

There is no correct size prescribed for a Group Discussion. The size of the group depends upon (a) the subject matter under discussion (b) the type of participants. Normally, 8 to 12 individuals in a group discussion panel is an ideal size. However, for highly professional and articulate people, the ideal size is 5/6 participants in the group, as the participants would have more to contribute to the topic under discussion.

The group members should be such that it reflects the characteristics of a particular market segment under study. The sampling plan is drawn up first, which then helps in deciding the composition of the group discussion panel members.

The role of the moderator in a group discussion is to establish a rapport with the group members. He/she also informs the members of the rules to be followed in the group discussion. The moderator acts as a stimulus to provoke heated discussion in the group, so that all underlying ideas, feelings and emotions emerge. The moderator also intervenes after time intervals, to asses the extent of agreement/disagreement with the ideas that are emerging.

Application areas for Group Discussion

The following are the situations where a group discussion could be used:

1) For development of new product ideas, abstract concepts.

2) For product positioning ideas — how does to consumer perceive a particular product/service/idea.

3) To determine hidden feelings/attitudes of the consumer.

4) To establish consumer vocabulary, as a preliminary step when developing questionnaires. It is not always possible to think of all questions associated with a product/brands, when conducting survey research. The group discussion helps in developing suitable questions.

5) For advertising and communication research - what to say to the consumer, how to say it, what should be the focus of the message etc.

Advantages of a Group Discussion:

1) In a Group Discussion setting, the interaction among members acts as a stimulant to generate new ideas, which may never be possible in an in-depth individual interview situation.

2) The Group Discussion setting leads to spontaneity in conversation which cannot come in individual interview situation.

3) The Group setting is emotionally provocative - the conversation may set off the thinking process, and one would recall older memories and conversations that may be relevant to the discussion on hand.

Disadvantages of a Group Discussion:

1) Qualified and trained personnel are required to conduct Group Discussion.
2) Analysis and interoperation is highly subjective.
3) A few members in a group could dominate the entire discussion.

Activity 2

Conduct a focus group interview with six people in your organization on the purchase and consumption of (a) fast food (b) Packed fruit juices. Prepare a report based on your result

.....

.....

.....

.....

3. Projective Techniques

When a researcher is conducting an in-depth interview, or conducting a survey through the questionnaire method, he might face problems in the form of language barriers with the respondent, or illiterate respondent (especially in social research and rural research), or social barriers (respondent is embarrassed to talk about a topic) or psychological barriers (recall of event or feelings at that moment of interview is not there or trying to avoid certain questions or "can't say" answer). In order to overcome such barriers faced during an interview process, the researcher may replace the questionnaire with the projective techniques.

The basic underlying concept behind projective techniques is that in certain situations, it is impossible to obtain correct information about what a person thinks or feels by asking him to describe his/her feelings - but this information can be obtained by making the respondent to project his feelings on to some other person or object.

The following are the different types of projective techniques:

i) **Word Association Test :**

In this method, the respondent is presented with a list of stimulus words, and for each word, is asked to respond with what he thinks about the word. The respondent is not given time to think of the responses. The ideas is that the 'first thought' responses are likely to reveal the true feelings of the respondent about the stimulus.

Example: Suppose a company wishes to understand what a person thinks about leisure time:

Stimulus word	Response
• Train	
• Books	
• Television	
• Swimming	
• Automobile	
• House	

The first responses to these words would give a fair idea as to how the respondent would like to spend his leisure time.

The analysis of the word association test is done as follows:

- Determine the frequency with which the various responses emerge.
- Category of attitudes can be created (e.g., favourable/neutral/unfavourable), and words.

Example:
If one wants to study the perception of students regarding MBA course, the categories that could emerge are:

practical	impractical
interesting	boring
easy	difficult
Job oriented	theory oriented.

The words that emerge from word association tests can be useful in developing semantic differential scale, likert scale, adjectives that describe products/services.

ii) **Sentence Completion Test:**

This is an extension of the word association test. In this method, the respondent is asked to finish an incomplete sentence with the first thought that comes to his mind. The idea is that the respondent projects his own feelings into the sentence.

Example:

- I like to drive mobikes because.....
- Most people smoke because.....
- I smoke because.....
- Cars are .....

As with the word association, the frequency of responses are taken into consideration while analyzing the results of the sentence association test.

iii) Fantasy Situation:

Here, the respondents are asked to imagine that they are converted into a product itself e.g., car, box of chocolate. This leads to the respondent imagining himself to be product itself and give the human characteristics to the product. This method is used for developing brand perception, brand personality.

iv) Cartoon Completion:

In this method the respondent is shown a cartoon that is similar to a comic strip, with "balloons" indicating speech. Usually, two people are shown talking to each other about a particular product/service/situation, but only one balloon contains the speech.

The situation that is shown in the cartoon is obviously of special interest to the researcher, and is part of the research project under hand. The respondent has to fill the other 'balloon' with his answer to what the other person is saying.

With this one tries to measure attitude towards a product or service. Analysis and interpretation of these results are highly subjective.

v) Picture Interpretation (Thematic Apperception Test) :

Thematic Apperception Test (TAT), along with the Rorschach Inkblot test, is probably the most widely known and used projective technique in Clinical Psychology. The same basic technique used above is applied for marketing research applications.

Here, the respondent is shown a picture - either a line drawing, illustration or photograph which is rather ambiguous, and is asked to describe what is going on - or tell a story about what is illustrated.

For example, a retail outlet like a gift shop is interested in knowing what a shopper thinks while buying a gift, how does he look at the various display items, what he expects during the shopping experience - for such a research project, a researcher might show the photograph of a person entering the gift shop to the respondent, and ask him to develop a story on what he thinks could be happening inside the shop. Normally, a respondent would project his own feelings while developing the story from the photograph. This gives an idea to the researcher as to what the respondent expects on entering the gift shop. The major drawback with this method is that there is a great deal of subjectivity in interpretation of the responses to these projective tests. These projective techniques are used in exploratory research, whose output act as input for developing hypothesis and full-fledged quantitative research.

Activity 3

Select any poduct and brand of interest to you. Administer each of the following techniques to five fellow colleagues to develop an idea of their feelings towards the brand:(a) Word Association Test, (b) Sentence Completion Test, (c) Cartoon (d) Thematic Apperception Test

.....

.....

.....

.....

4. Observation Method

Observation method is another very powerful tool for getting information about the consumer. This method is used for recording behaviour of people, objects, events. Informal observations are extensively used for observing customer buying patterns, impact of

competitive advertisement on buying, product availability etc.

Observation technique is always used in conjunction with other research techniques. The inherent danger in this method is that one could draw wrong conclusions as it is highly subjective, and a lot depends upon the observer's perception of a situation.

Advantages of Observation Technique:

1) This method does not rely on the willingness of the respondent to cooperate and provide information.

2) Behaviour patterns of which a respondent is not aware of, can be recorded by observation only e.g., facial expression of a consumer while examining a new product display in a store.

Disadvantages:

1) By observation method one cannot observe a consumers beliefs, feelings, awareness etc (other research techniques like focus groups, in-depth interviews required).
2) The observed behavior pattern must be of short duration, should occur frequently, in order to qualify for observation, and act as input to other research techniques.

Classification of Observation Techniques:

a) Natural vs. Contrived Observation

Natural observation is recording of consumer behaviour as it takes place e.g., shopping process. Contrived observation is one in which an artificial environment is created and then consumer behaviour is observed e.g., Simulated Test Marketing.

b) Disguised vs. Undisguised Observation

Disguised refers to whether or not consumer is aware of being observed. e.g., two-way mirrors, hidden cameras in shops, observers dressed as sales clerks etc.

c) Structured vs. Unstructured Observation

In a structured observation, the decision problem has been clearly defined. Information needs are clearly defined, and so this reduces observer bias and increases the reliability of observed data.

In an unstructured observation, the problem is yet to be defined and formulated, and is ideal for exploratory research.

d) Human vs. Mechanical Observation

By replacing humans with mechanical devices, the accuracy of observation increases, observer bias is reduced and lower costs are incurred. Special mechanical devices used for observation could be video cameras (shopping behaviour), Audimeter/Peoplemeter (TRP ratings), Eye-cameras, etc.

Activity 4

Observe few shoppers of chocolates in your neighbourhood store. What insights, you could gain from these observations regarding purchase of chocolates.

.....

.....

.....

.....

8.6 LATEST TECHNIQUES OF QUALITATIVE RESEARCH

Qualitative Research is taking more and more inputs from psychology, to develop latest techniques in Qualitative Research. Some of the latest techniques used are as follows:

1) **Imagery Technique:** This technique is used to determine the images of two competing brands. As an example, reades were asked to describe two competing magazines.
Magazine 1 was described as young, attractive, bubbling teenager.
Magazine 2 was described as older, less interesting, introvert lady. This information helped the owner of the magazine to reposition magazine 2.

2) **Synectics** This technique gives the researcher an insight into the intuitive or emotional impressions of the consumer for either a brand or advertising campaign by unleashing the creative side of the consumer. The respondents are asked to go through a series of games designed to lower their inhibitions which help in unleashing the child present in every adult. The basic idea is that creativity works at child level. With this exercise, the researcher can get original ideas for use in developing advertising campaign, as also the consumer's real feelings about products.

3) Neuro - Linguistic Programming : This technique provides an insight into the 'Non-Verbal imapact of an adevertisement on the target consumer. This is done by analyzing as to which of the five senses (smell. taste touch, sight, aural) a consumer uses to recall an advertiisement. As an example, the Onida advertisement campaigh was analyzed using this technique, and it was able to reveal strong sensory effect of Rasping Voice and Shattering Glass. The underlying idea behind this technique is that much of the recording of experience goes on below the surface' and outside conscious awareness. For the purposes of marketing. there is a condification of the sersory modalities into Visual. Auditory and Kinesthetic (touch,taste and smell The shorthand version is called as VAK segmentation. Human are emotiona 'creatures, and a person would buy on emotion and justify the purchase on 'fact Thus, to decode the emotions surrounding the product purchase and experience researchers use the NLP technique. NLP techniques are modelled on techniques used in other fields of psychology (EricksonianHypnosis and Classical Conditioning), and hence only persons trained in the field of psychology can use this technique

8.7 SUMMARY

The variours methods by which infomation about consumers can obtained are broadly classified into two categories. Qualitative Research and Quantitative Research. Qualitative Research may be used to generate hypothesis for further research, is exploratory in nature and involves a small number of people who are not sampled on probabilistic basis. Qualitative research. can provide leads and insights which could be of extreme help in conducting quantitative research. There are various methods of conducting qualitative research. These are depth interviews, focus groups, observation and projective techniques like word association test, sentence completion test, coroon thematic apperception test. Qualitative research is increasingly using inputs from psychology and some of latest techniques being used include imagery, synectics and neuro- linguistic programming

8.8 SELF-ASSESSMENT QUESTIONS

1) Highlight the main differences between qualitative and quantiresearch

2) Wich the help of examples, discuss the areas where qualitative research can be used in marketing

3) What are the various methods of conducting qualitative research? Discuss advantages and limitations of each of these methods.

8.9 FURTHER READINGS

Parasuraman, A *" Marketing Reseach", Addision* - Wesley Publishing Company Inc., USA.

Chisnall, P. *" The Essence of Marketing Research"*, PHI, New Delhi

Tull, D. And Hawkins, D, *"Marketing Research - Measurement and Method"*. Macmillan Publishing Co., New York

UNIT 9 DATA PROCESSING - CODING, TABULATION AND DATA PRESENTATION

Objectives

After studying this unit you should be able to :

- distinguish between two types of editing
- design the coding scheme for a questionnaire
- classify the data according to attributes or numerical characteristics and
- differentiate between simple and manifold classification of data
- present the data in the forms of tables and graphs
- put the data in the form of frequency distribution

Structure

9.1 Introduction
9.2 Editing of Data
9.3 Coding of Data
9.4 Classification of Data
9.5 Statistical Series
9.6 Tables as Data Presentation Devices
9.7 Graphical Presentation of Data
9.8 Summary
9.9 Self -Assessment Questions
9.10 Further Readings

9.1 INTRODUCTION

The survey data collected from the field should be processed and analyzed as indicated in the research plan. Data processing primarily involves editing, coding, classification and tabulation of data, so that it becomes amenable for data analysis. This unit concentrates on various aspects of data processing. The processing of data can either be in the form of tables or in the form of graphs. These aspects have been widely covered in unit 6, block 2 of the course on Quantitative Analysis for Managerial Applications (MS-8). You are therefore advised to go through it before reading this unit.

9.2 EDITING OF DATA

The editing of data is a process of examining the raw data to detect errors and omissions and to correct them, if possible, so as to ensure legibility, completeness, consistency and accuracy.

The recorded data must be legible so that it could be coded later. An illegible response may be corrected by getting in touch with people who recorded it or alternatively it may be inferred from other parts of the question.

Completeness involves that all the items in the questionnaire must be fully completed. If some questions are not answered, the interviewers may be contacted to find out whether he failed to respond to the question or the respondent refused to answer the question. In case of former, it is quite likely that the interviewer will not remember the answer. In such a case the respondent may be contacted again or alternatively this particular piece of data may be treated as missing data.

It is very important to check whether or not respondent is consistent in answering the questions. For example there could a respondent claiming that he makes purchases by credit card may not have one.

The inaccuracy of the survey data may be due to interviewer bias or cheating. One way of spotting is to look for a common pattern of responses in the instrument of a paritcular interviewer.

Apart from ensuring quality data this will also facilitate in coding and tabulation of data. In fact, the editing involves a careful scrutiny of the completed questionnaires

The editing can be done at two stages:

1. Field Editing, and
2. Central Editing.

Field Editing : The field editing consists of review of the reporting forms by the investingator for compieting or translating what the latter has written in abbreviated form at the time of interviewing the respondent. This form of editing is necessary in view of the writing of individuals, which vary from individual to individual and sometimes difficult for the tabulator to understand. This sort of editing should be done as soon as possible after the interview, as it may be necessary sometimes to recall the memory. While doing so, care shold be taken so that the investigator does not correct the errors of omission by simply guessing when the respondent would have answered if the question was put to him.

Central Editing: Central editing should be carried out when all the forms of schedules have been completed and returned to the headquarters. This type of editing requires that all the forms are thoroughly edited by a single person (editor) in a small field study or a small group of persons in case of a large field study. The editor may correct the obvious errors, such as an entry in a wrong place, entry recorded in daily terms whereas it should have been recorded in weeks/months, etc. Sometimes, inappropriate or missing replies can also be recorded by the editor by reviewing the other information recorded in the schedule. If necessary, the respondent may be contacted for clarification. All the incorrect replies, which are quite obvious, must be deleted from the schedules.

The editor should be familiar with the instructions and the codes given to the interviewers while editing. The new (corrected) entry made by the editor should be in some distinctive form and they be initialed by the editor. The date of editing may also be recorded on the schedule for any future references.

Activity 1

Define the following.

a) Field Editing.

.....
.....
.....
.....

b) Central Editing.

.....
.....
.....
.....

Activity 2

A marketing research organization is conducting a survey to determine the consumption pattern of food items by households in Delhi. You are the head of computer division responsible for editing the raw data from the questionnaires and analyzing the same. A filled up set of questionnaires have been sent to you. List out the points on which you would like to concentrate while editing the raw data.

9.3 CODING OF DATA

Coding is the process of assigning some symbols (either) alphabetical or numerals or (both) to the answers so that the responses can be recorded into a limited number of classes or categories. The classes should be appropriate to the research problem being studied. They must be exhaustive and must be mutually exclusive so that the answer can be placed in one and only one cell in a given category. Further, every class must be defined in terms of only one concept.

The coding is necessary for the efficient analysis of data. The coding decisions should usually be taken at the designing stage of the questionnaire itself so that the likely responses to questions are pre-coded. This simplifies computer tabulation of the data for further analysis. It may be noted that any errors in coding should be eliminated altogether or at least be reduced to the minimum possible level.

Coding for an open-ended question is more tedious than the closed ended question. For a closed ended or structured question, the coding scheme is very simple and designed prior to the field work. For example, consider the following question.

- **What is your sex?**

Male ☐ Female ☐

We may assign a code of '0' to male and '1' to female respondent. These codes may be specified prior to the field work and if the codes are written on all questions of a questionnaire, it is said to be wholly precoded.

The same approach could also be used for coding numeric data that either are not be coded into categories or have had their relevant categories specified. For example,

- **What is your monthly income?**

Here the respondent would indicate his monthly income which may be entered in the relevant column. The same question may also be asked like this:

- **What is your monthly income?**
  - < Rs. 5000
  - Rs. 5000 - 8999
  - Rs. 13000 - 12999
  - Rs. 13000 or above.

We may code the class 'less than Rs.5000' as '1', Rs. 5000 - 8999' as '2', 'Rs. 9000 - 12999' as '3' and 'Rs. 13000 or above' as '4'.

Coding of open-ended questions is a more complex task as the verbatism responses of the respondents are recorded by the interviewer. In what categories should these responses be put to? The researcher may select at random 60-70 of the responses to a question and list them. After examining the list, a decision is taken to what categories are appropriate to summarize the data and the coding scheme for categorized data as discussed above is used-A word of caution-that while classifying the data into various categories we should keep provision for "any other" to include responses which may not fall into our designated categories.

It may be kept in mind that the response categories must be mutually exclusive and collectively exhaustive.

A study was carried out among the readers of newspapers with the following objectives.

- To identify and understand the factors that determine the preference for Times of India amongst the readers.
- To identify the profile of the readers of Times of India.
- To ascertain the expectations vs. perceptual reality and locate gaps if any amongst the readers of Times of India.
- To analyze the factors responsible for the most preferred subjects of information attracting the readers to prefer Times of India.

To achieve these objectives a questionnaire was designed. We give below a part of the questionnaire, and discuss the coding scheme for the same. Please note that the objective here is not to evaluate the questionnaire but to design the coding scheme for any given questionnaire of a study. The said questionnaire is given below in Exhibit 1.

Exhibit - 1

1. *Which of the following newspapers do you read?*

   a) *The Times of India*
   b) *The Hindustan Times*
   c) *Indian Express*
   d) *Pioneer*
   e) *Any other (Please specify)*

If the respondent does not read Times of India, move to question number 6.

2. *I have been subscribing to TOI for .... (please tick as applicable)*

   a) *Less than 6 months* _____
   b) *6 months - less than 1 year* _____
   c) *1 year - less than 2 years* _____
   d) *Two years or more* _____

3. *Please rate the following features on a scale of 1 to 5, as per the interest you have in each one of them.*
   (5 - Very interested, 4 - Interested, 3 - Not particularly, 2 - Uninterested, 1 - Very uninterested)

   a) *Political News* _____
   b) *International News* _____
   c) *City News* _____
   d) *Corporate and Business News* _____
   e) *Sports News* _____
   f) *People and Lifestyles News* _____
   g) *Leisure, Art and Entertainment News* _____

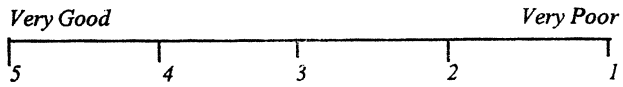
4. *Please rate TOI on the following attributes along the scale marked below*

   a) *News Content*

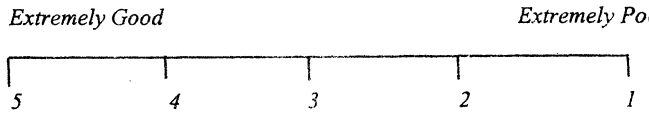
   Very Adequate ——— *Very Inadequate*

   5 4 3 2 1

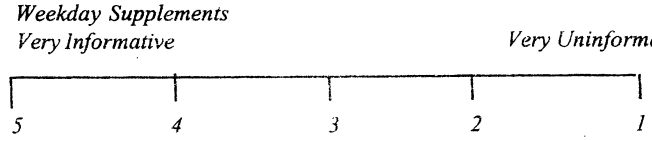
b) *Editorial*



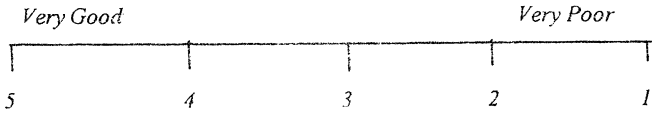
c) *Weekend Supplements*



d) *Weekday Supplements*



e) *Layout*



5. *How would you rank the following features from 1 to 7 in terms of importance to you? (1 indicates very important - - - - - - - - - - - - - - 7 indicates least important)*

a) *Classified Advertisements* ____
b) *Weekday Supplements* ____
c) *Number of pages in the issue* ____
d) *Advertisements* ____
e) *News Content* ____
f) *Weekend Supplements* ____
g) *Layout* ____

6. *Please fill out the following details:*

a) *Age :*
b) *Sex : Male Female*
c) *Total number of members in the household* ____
d) *Occupation : Business*
Service
Any other (pl. specify) : ____
e) *Monthly Income of the household (in Rupees)*
Less than or equal to 5000 ____
5001 - 10000 ____
10001 - 15000 ____
15001 - 20000 ____
20001 - 25000 ____
Above 25000 ____

- Let us design the coding scheme for the questionnaire given in exhibit 1. We note that question number 1 may have multiple responses because a respondent could read one or more than one newspapers. There are 5 alternatives assigned for question number 1 and therefore we will use five columns in the data matrix to record the responses of this question. If the respondent reads Times of India we code it a value 1 otherwise 0. Similarly it is done for the remaining newspapers. However, if there is a respondent who read Times of India and Indian Express we will code question 1a and 1c having a value of 1 and for the remaining parts namely b, d and e the coded value will be 0.

For question number 2, the respondent can choose only one of the four alternatives. Therefore one single column is required to record the responses of the respondents. The response categories are mutually exclusive and collectively exhaustive. Whichever category is chosen by the respondent that is coded 1 and the remaining are coded 0.

- Question number 3 has seven parts and the respondent is to rate each one of them on a 5-point scale ranging from 1 to 5. Therefore a total of seven column is required to record the responses of the respondent. Suppose the respondent rates International News as 4 the value of 4 should be assigned to question number 3b and so on.

- There are five attributes of Times of India mentioned in question number 4 and the respondent is assigned the job of rating each of them on a scale of 1 to 5. Therefore five columns are required to record the responses of this question. Suppose for question 4c (Weekend Supplements) the rating of the respondent is 2, and the same will be shown in the coding book corresponding to this question.

- There are six features of Times of India mentioned in question number 5 and labeled as 5a to 5f. The respondent is to rank them from 1 to 6 with regards to the importance it gives to each of these features. Therefore we need six columns for this. Suppose the rating is 2, 3, 6, 1, 4 & 5 for questions numbering 5a to 5f respectively. The same numbers would appear on the coding sheet corresponding to this question.

- Question number 6 is divided into five parts. For each of the part one separate column is required. 6a indicates the age of the respondent which will be indicated as per the data revealed by the respondent. Question 6b is concerning the sex of the respondent. Here male respondents are coded as 1 whereas female respondents are coded as 0. Question 6c indicates the total number of members in the household. Question 6d is concerned with the occupation of the respondent. Question 6e mention the monthly income of the household put in categorized form. Here the responses are mutually exclusive and collectively exhaustive. If the respondent has a monthly income of less than 5000 rupees, the response is coded as 1, if monthly income is between 5001-10000 rupees, it is coded as 2, in case it is between 10001-15000; the code is 3. From 15001-20000: the code is 4. 20001-25000; the code is 5 and above 25000; the code is 6. The above discussion can be shown below in the form of a code book.

CODE BOOK FOR THE SAMPLE QUESTIONNAIRE

Q.No.	No. of Columns required	Variable Name	Category Definition
1.	5	News Paper	TOI = 1, otherwise 0 HT = 1, otherwise 0 IE = 1, otherwise 0 P = 1, otherwise 0 AO = 1, otherwise 0
2.	1	Duration	< 6 months = 1 6 - 12 months = 2 12 - 24 months = 3 24 or more = 4
3.	7	Rating of Features	Each of the seven variable can take any value between 1 to 5.
4.	5	Attribute	Each of questions 4a to 4e can take any value between 1 to 5.
5.	6	Ranking of Feature	Each of the feature could take any value between 1 to 7.
6a.	1	Age	Two digited numbers.
6b.	1	Sex	Male = 1 Female = 0
6c.	1	Numbers in household	Two digited numbers.
6d.	1	Occupation	Business = 1 Service = 2 Any other = 3
6e.	1	Income	≤ 5000 = 1 5001 - 10000 = 2 10001 - 15000 = 3 15001 - 20000 = 4 20001 - 25000 = 5 Above 25000 = 6

The data matrix corresponding to the above coding scheme is shown in the table given below:

DATA MATRIX FOR THE CODED SAMPLE QUESTIONNAIRE

Rsp	Q 1a	Q 1b	Q 1c	Q 1d	Q 1e	Q 2	Q 3a	Q 3b	Q 3c	Q 3d	Q 3e	Q 3f	Q 3g	Q 4a	Q 4b	Q 4c	Q 4d	Q 4e	Q 5a	Q 5b	Q 5c	Q 5d	Q 5e	Q 5f	Q 5g	Q 6a	Q 6b	Q 6c	Q 6d	Q 6e
1	1	0	1	0	0	2	1	4	3	5	1	4	4	4	3	5	3	5	7	4	6	3	1	2	5	32	1	4	1	3
2	0	1	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	30	0	3	2	4
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
nth	1	1	0	0	0	3	2	5	1	4	2	5	5	4	5	4	2	4	6	5	7	1	4	3	1	41	1	5	2	5

The above table indicates that the respondent number 1 reads both Times of India and Indian Express and no other newspaper. This is indicated by code 1 corresponding to question 1a and 1c and for the remaining parts of questions 1 a '0' is indicated. Question number 2 indicates that the respondent is reading Times of India from 6 to 12 months. The rating of various features of a newspaper in terms of the interest he has in them is indicated by responses indicated in questions 3a to 3g. The respondent is not very uninterested in critical news, interested in international news, not particular about city news, very interested in corporate and business news, very uninterested in sports news and interested in people and lifestyle news; and leisure art and entertainment news. The respondent rates Times of India on five attributes. He can give a possible rating of Times of India on various attributes on a scale of 1 to 5 where 1 is on extremely unfavorable side whereas 5 represents extremely favourable side. He has rated Times of India on news content as 4, editorial as 3, weekend supplements as 5, weekdays supplements as 3 and layout as 5. However, his ranking of how various features are important to him on 1 to 7 scale, where 1 represents very important and 7 the least important is indicated in question 5. As per the respondent, classified advertisements are ranked the least, weekdays supplements get a rank of 4, number of pages get a rank of 6, advertisement; a rank of 3, news content; a rank of 1, weekend supplements; a rank of 2, and layout; a rank of 5. The respondent is of 32 years of age and is a male as indicated by a code of 1 to question 6b. There are four members in his household. His occupation is business and has a monthly income between Rs. 10001 to 15000 as indicated by code 3 for question 6c.

Respondent 2 does not read Times of India. In fact the respondent is a reader of Hindustan Times and no other newspaper and therefore the questions mentioned in questions numbering 6 are asked to the respondent. The respondent is 30 years of age, and a female as indicated by code 0 for question 6b. The respondent has 3 family members, is a professional and have monthly income between 15001 - 20000 rupees as indicated by code 4 corresponding to 6c.

Activity 3

Describe the characteristics of nth respondent as given in data matrix ..

.....

.....

.....

.....

9.4 CLASSIFICATION OF DATA

In most research studies, voluminous raw data collected through a survey need to be reduced into homogeneous groups for any meaningful analysis. This necessitates classification of data, which in simple terms is the process of arranging data in groups or classes on the basis of some characteristics. Classification condenses the data, facilitates comparisons, helps to study the relationships and facilitates in statistical treatment of data. The classification should be unambiguous and mutually exclusive and collectively exhaustive. Further, it should not only be flexible but also suitable for the purpose for which it is sought. Classification can either be according to attributes or numerical characteristics.

1. **Classification According to Attributes:** To classify the data according to attributes we use descriptive characteristics like sex, caste, education, user of a product etc The descriptive characters are the one which can not be measured quantitatively. One can only talk in terms of its presence or absence. The classification according to attributes may be of two types.

   i) **Simple Classification**: In the case of simple classification each class is divided into two sub classes and only one attribute is studied viz. user of a product or non-user of a product, married or unmarried, employed or unemployed, Brahmin or non-Brahmin etc.

   ii) **Manifold Classification:** In the case of manifold classification more than one attributes are considered. For example, the respondents in a survey may be classified as user of a particular brand of a product and non-user of particular brand of product. Both user and non-user can be further classified into male and female. Further one can classify male and female into

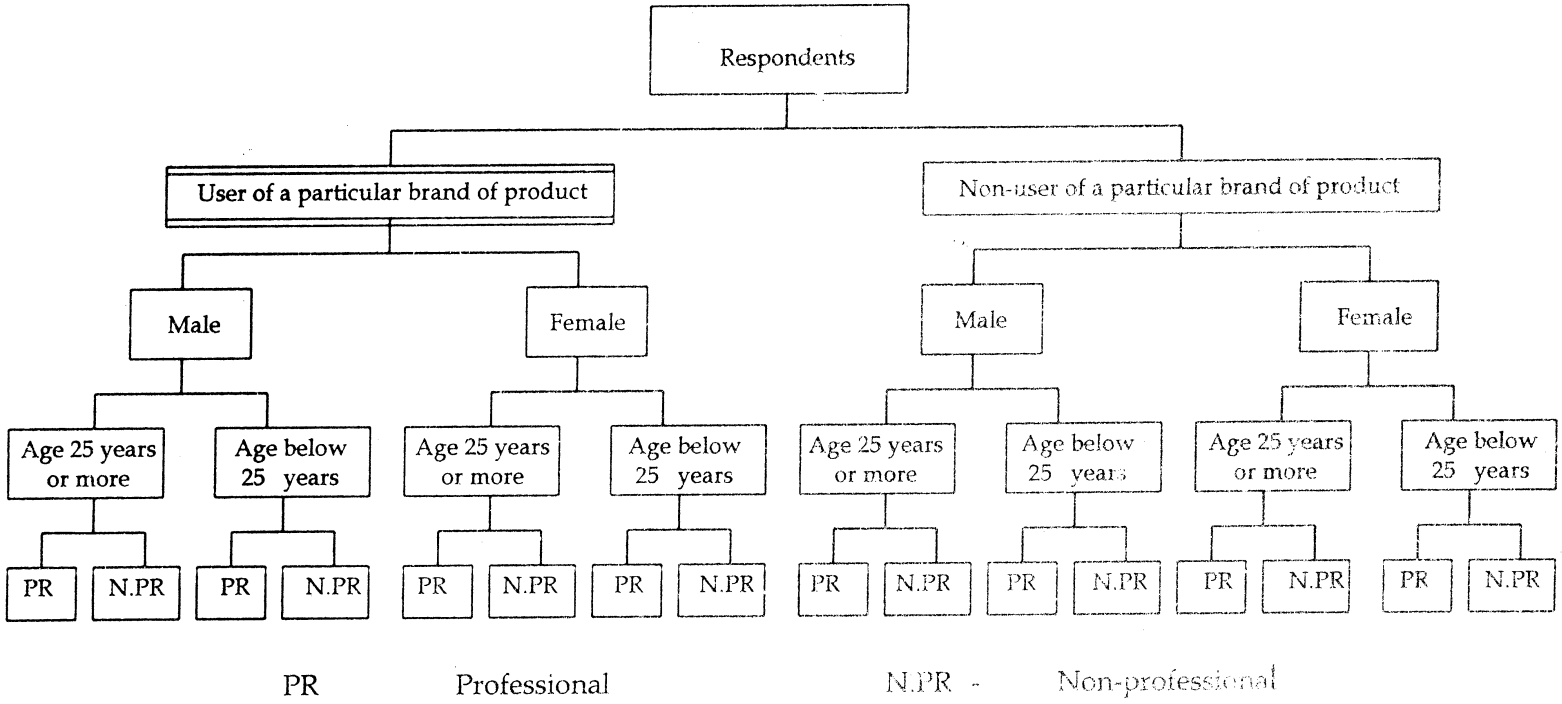


FIGURE - 1 : An Example of Manifold Classification

two categories such as below 25 years of age and 25 and more years of age. We can further classify them as professionals or non-professionals. This way one can keep on adding more attributes. This is shown in Figure - 1. However, the addition of a particular attribute (process of sub-classification) depends upon the basic purpose for which the classification is required. The objectives of such a classification has to be clearly spelt out.

Classification According to Numerical Characteristic: When the observations possesses numerical characteristics such as sales, profits, height, weight, income, marks, they are classified according to class intervals. For example, persons whose monthly income is between Rs. 2001 and Rs. 3500 may form one group, those whose income is within Rs. 3501 and Rs. 7000 may form another group, and so on. In this manner, the entire data may be divided into a number of groups or classes, which are usually called class-intervals. The number of items in each class is called the frequency of the class. Every class has two limits: an upper limit and a lower limit, which are known as class limits. The difference between these two limits is called the magnitude of the class or the width of the class interval. The class intervals may be formed by using inclusive and exclusive method. Suppose we have the class intervals such as 10 - 15, 16 - 21, 22 - 27 etc. Such a class interval is an example of inclusive method because both the lower and upper limit are included in the class. If the variable X falls in the first class interval, it can take values like $10 <= X >= 15$. The class intervals like 10 - 15, 15 - 20, 20 - 25 etc. form an example of exclusive class interval since the lower limit is included whereas the upper limit is excluded from the class interval. The variable X if falling in the first class interval, would take values as $10 <= X < 15$. As an illustration of how the data can be classified into class intervals using inclusive and exclusive method, we may consider the following example.

Example: Following data refers to the sales of a company for the 40 quarters. Tabulate the data using the inclusive method.

Qtr.	Sales	Qtr.	Sales	Qtr.	Sales	Qtr.	Sales
1	1060	11	1255	21	1690	31	1200
2	2125	12	1190	22	2130	32	2190
3	1440	13	870	23	1870	33	1800
4	1940	14	1460	24	1875	34	2255
5	2060	15	2125	25	1650	35	2000
6	1310	16	750	26	945	36	1060
7	2120	17	1120	27	2240	37	1370
8	2560	18	2000	28	1700	38	2375
9	2250	19	1750	29	1165	39	1470
10	2135	20	1760	30	1945	40	2250

We will be using the data given above. We form five class intervals each of width 370. These are inclusive class intervals in the sense that the variable X could take any value between the lower and upper limit in such a way that both ends of the interval could be covered under this. The class intervals alongwith the number of items in each class interval is shown in the table below:

Class Intervals	Tally Marks	Frequencies
750-1120	𝍸 |	6
1121-1491	𝍸 𝍸	9
1492-1862	𝍸 |	6
1863-2233	𝍸 𝍸 𝍸	13
2234-2604	𝍸 |	6

Activity 4

A survey was conducted to estimate the expenditure of households on entertainment. The data on a number of variable was collected. One of the variable of interest being monthly .income. A sample of 60 respondents gave the following data on monthly household income Use the data to form class interval using exclusive method

Rsp	Inc	Rsp	Inc	Rsp	Inc	Rsp	Inc	Rsp	Inc	Rsp	Inc
1	4560	11	4200	21	6700	31	11500	41	10000	51	7100
2	9824	12	5432	22	11660	32	7342	42	4660	52	6200
3	7336	13	98570	23	4050	33	8956	43	9876	53	65345
4	5874	14	10560	24	7890	34	4322	44	5656	54	5768
5	12000	15	12000	25	8845	35	6798	45	8888	55	7880
6	10660	16	10095	26	4560	36	4789	46	6999	56	4665
7	11650	17	6790	27	4824	37	5432	47	7645	57	5668
8	4465	18	7780	28	7366	38	98570	48	7891	58	8655
9	5378	19	5430	29	4874	39	10560	49	9800	59	5475
10	8765	20	4400	30	11200	40	9000	50	4050	60	11671

.....

.....

.....

.....

9.5 STATISTICAL SERIES

A series is defined as a logical or systematic arrangement of observations or items. When the attributes or things are counted, measured or weighed and arranged in an orderly manner, say either descending or ascending order, they constitute a series. When the statistical data pertains to time, the series is said to be historical or time series. The important factor in such series is the chronology. In time series data, the time difference between any two observations must be same. It could be hour, minute, week, month, quarter, year etc. The data presented in following table on sales forms a time series data.

Year	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Sales (Rs. lacs)	60	75	68	74	82	73	91	96	90	102

When the data pertains to space, the series is referred to as special, and is also known as geographical series. When the data refers to physical conditions such as height, weight, age, etc., the series is referred to as condition series. The following series, for example, refers to special series of regionwise sales of a firm during 1989-90.

Region	Sales (Rs. in lakhs)
Northern	135.00
North-eastern	56.00
Southern	85.00
Central	76.00
Western	163.00
Eastern	68.00

The series can also be classified as individual observations, discrete series and continuous series. In case of series of individual observations, the items are listed singly as distinguished from listing them in groups. In case of discrete series, items are arranged in groups (frequency distribution) showing definite breaks from one point to another and are exactly measurable. The grades obtained by 70 students in Marketing Research are given in the following frequency distribution table.

Grades	F	D-	D	D+	C	C	C+	B-	B	B+	A-	A	A+
No. of Students	2	6	5	4	7	10	12	6	7	5	3	4	1

In case of continuous series, the items are arranged in class and they can be arranged either in ascending order or descending order of magnitude and their continuity is not broken. At the point at which a class ends, the next begins and thus the continuity is maintained. The distribution of the lifetime of 350 radio tubes is given below in the form of a continuous series.

Life-Time (in-hours)	No. of tubes with life time	Life-time (in hours)	No. of tubes with life time
300-400	6	700-800	62
400-500	18	800-900	22
500-600	73	900-1000	4
600-700	165		

Activity 5

Collect the data on the salaries of the employees in your organisation and develop a continous series for the data you have collected

.....

.....

.....

.....

Activity 6

State what lebels you will attach to the following series of data

i) **Density of population (per sq. km in different cities of India)**

.....

ii) **Number of defective items produced by machine per week**

.....

iii) **Population of workers classified as male, female and further classification is based on their being vegetarian or non-vegetarian**

.....

9.6 TABLES AS DATA PRESENTATION DEVICES

Statistical data can be presented in the form of tables and graphs. In the tabular form, the classification of data is made with reference to time or some other variables. The graphs are used as a visual form of presentation of data.

The tabulation is used for summarization and condensation of data. It aids in analysis of relationships, trends and other summarization of the given data. The tabulation may be simple or complex. Simple tabulation results in one-way tables, which can be used to answer questions related to one characteristic of the data. The complex tabulation usually results in two way tables, which give information about two interrelated characteristics of the data; three way tables which give information about three interrelated characteristics of data; and still higher order tables, which supply information about several interrelated characteristics of data.

Following are the important characteristics of a table:

i) Every table should have a clear and concise title to make it understandable without reference to the text. This title should always be just above the body of the table.
ii) Every table should be given a distinct number to facilitate easy reference.
iii) Every table should have captions (column headings) and stubs (row headings) and they should be clear and brief
.iv) The units of measurements used must always be indicated.
v) Source or sources from where the data in the table have been obtained must be indicated at the bottom of the table.
vi) Explanatory footnotes, if any, concerning the table should be given beneath the table alongwith reference symbol.
vii) The columns in the tables may be numbered to facilitate reference.
viii) Abbreviations should be used to the minimum possible extent.
ix) The tables should be logical, clear, accurate and as simple as possible.
x) The arrangement of the data categories in a table may be a chronological, geographical, alphabetical or according to magnitude to facilitate comparison.
xi) Finally, the table must suit the needs and requirements of the research study.

Activity 7

Construct a table showing the number of letters sent by ordinary male, registered post and through courier in your organisation for the last 10 years. Also show the number of inland and overseas leetters

.....

.....

.....

.....

Activity 8

You are working in a company dealing with exports of readymade garments. Prepare a table showing the exports to various countries for the financial year 1997-98

.....

.....

.....

.....

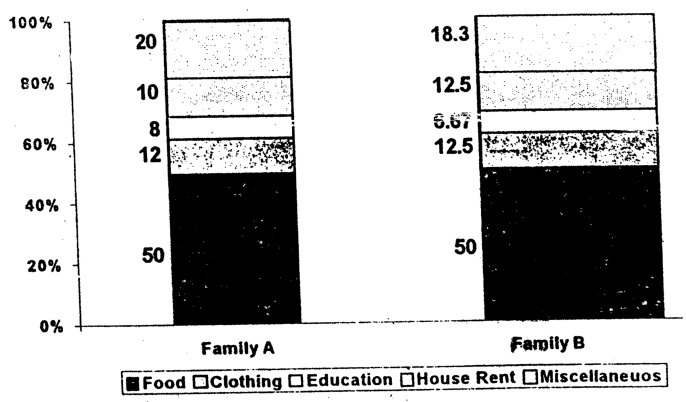
9.7 GRAPHICAL PRESENTATION OF DATA

Several types of graphs or charts are used to present statistical data. Of them, the following are commonly used: bar chart, two dimensional diagrams, pictograms, pie charts and arithmetic chart or line chart. Several of these have been discussed in detail in Unit 6, (Presentation of data), Block 2 (Data Collection and Analysis) of Quantitative Analysis of Managerial Applications (MS-8). You may therefore refer to the said study material. Therefore, charts and graphs not covered are discussed below.

1. **Two dimensional diagrams:** The commonly used two dimensional diagrams are rectangular diagrams and squares. In rectangular diagrams the rectangles are used to present the data in the graphic form. These diagrams are used for comparing two sets of data. The height of the rectangle is proportional to the ratio of the data which bear to each other in a given series and the width of the rectangle varies in proportion to the aggregate. Before constructing the rectangular diagram, the data is converted into percentages. A rectangular diagram to the data on household expenditures of various items by two families A and B is given below.

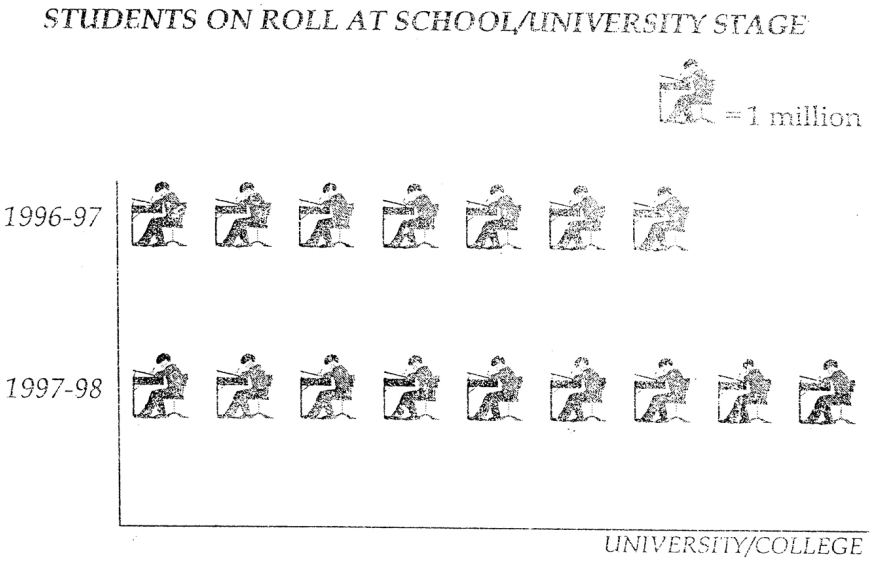
Items	Family A		Family B	
	Expenditure	%	*Expenditure*	%
Food	2000	50	3000	50
Clothing	480	12	750	12.5
Education	320	8	400	6.67
House Rent	400	10	750	12.5
Miscellaneous	800	20	1100	18.3
Total	4000	100	6000	100

The above data is presented below in two-dimensional diagram. In this diagram the width of the rectangles for family A to that of family B is taken in the proportion of 2 : 3



When the difference between two quantities is very large, one bar would become too big and the other too small in a rectangular diagram. To overcome this difficulty, squares are used to present the data. The size of the square is the square roots of the given data.

2. **Pictograms:** In this form of presentation, data are **representc by picture for** example, population figures are presented by the **Pivture of a human being** production figures of, say motorbikes, is presented by **the presented notorbike** cattle population by a picture of a cattle and so on. **The follwing and presents a** pictograph showing the students on roll at the **school/universal level toe the year** 1996-97 and 1997-98.



3. **Pie Chart:** In a pie chart, different segments of a circle represent percentage contribution of various components to the total. It brings out the relative importance of various components of data. For drawing a pie chart, we construct a circle of any diameter and then the circle is broken into a desired number of segments, angle 360^0 representing 100 percent. The expenditure on various commodities by two families A and B is given in the following table:

Monthly Expenditure on Various Commodities by Two Families (Figures in Rupees)

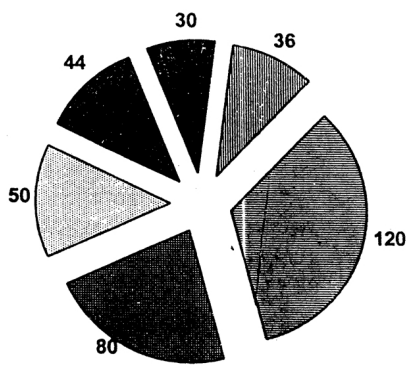
Type of Commodity	Family A	Family B
Food	3000	5000
Rent	2000	3500
Clothes	1250	250
Education	1100	225
Miscellaneous	750	1250
Savings	900	1500

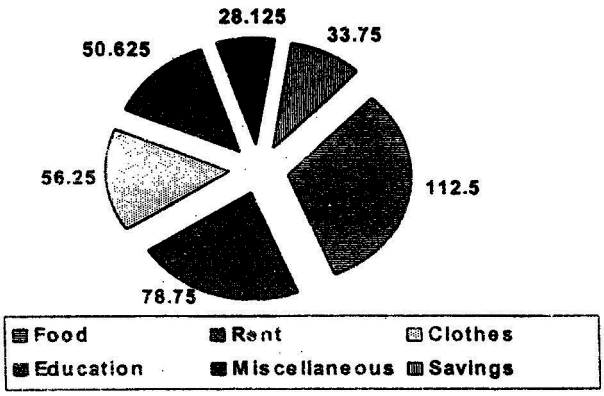
The various items are converted into corresponding degrees using the fact that the sum total of degrees in the circle equals 360. The degrees for various items corresponding to family A and B is given below:

Type of Commodity	Family A	Family B
Food	120.000	112.500
Rent	80.000	78.750
Clothes	50.000	56.250
Education	44.000	50.625
Miscellaneous	30.000	28.125
Savings	36.000	33.750
Total	360.000	360.000

The pie chart corresponding to the figures given in the above table is shown belo

FAMILY-A

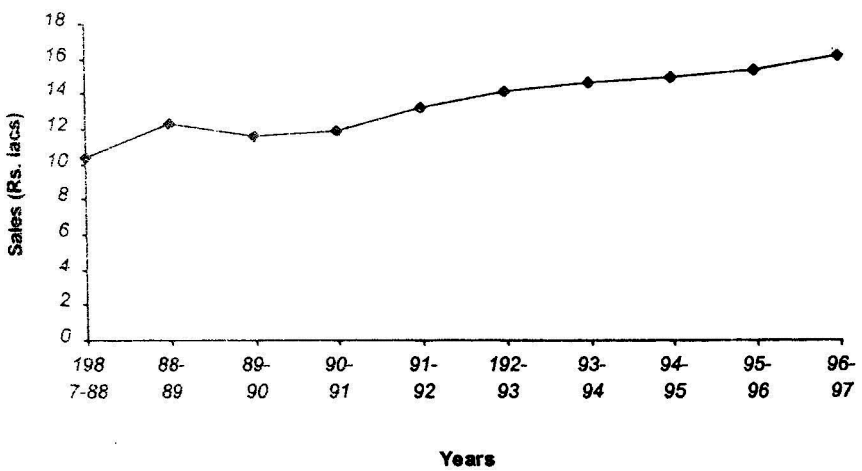




Line or Arithmetic Chart: The line or arithmetic chart is used to identify the changes or the trend that exist in a series of data. The data on export sales of a company between 1997-98 and 1996-97 is presented below in the form of a line chart. This data is given in the following table:

Years	*87-88*	*88-89*	*89-90*	*90-91*	*91-92*	*92-93*	*93-94*	*94-95*	*95-96*	*96-97*
Sales (Rs. lacs)	*10.4*	*12.3*	*11.6*	*11.9*	*13.2*	*14.1*	*14.6*	*14.9*	*15.4*	*16.2*

Although we can see changes in the data, the presentation of the same on a line chart gives a better picture of the information. The other forms of presentation of the data have already been discussed in Unit 6, Block 2 of Quantitative Analysis for Managerial Applications (MS-8).



9.8 SUMMARY

In this unit various aspects of data processing namely editing, coding, classification and presentation of data through tables and graphs have been discussed. Editing is of two type namely field editing and central editing. Coding involves assigning of some symbols or numerals or both to the answers of the questions in a questionnaire so that the responses can be recorded in a limited number of classes or categories. This helps in analysis of data. Designing of a coding scheme has been discussed with the help of a sample questionnaire. Classification is the process of arranging data in groups or classes on the basis of certain characteristics. It involves condensation of data which facilitates comparison and helps in establishing relationships between variables. Classification can be according to attributes or numerical characteristics. The former may be divided into simple and manifold classification. The later is achieved using either inclusive or exclusive method of forming frequency distribution. The data may be presented in the form of tables or graphs. The unit discusses various characteristics which should be taken into consideration while forming a table. The graphical presentation of data to be done by using pie chart, line chart, histograms etc. Some of these have been covered in this unit.

9.9 SELF-ASSESSMENT QUESTIONS

1. Describe, in brief, the importance of editing, coding, classification, tabulation and presentation of data in the context of research study.
2. Discuss the different aspects of classification of data. What are the likely problems encountered in the classification and how they can be handled?
3. Why tabulation is considered essential in a research study? Give the characteristics of a good table.
4. Write briefly about the different forms of data presentation devices.
5. Monthly salary of 32 employees of a firm are given below. Tabulate the data after selecting a suitable class interval.

2250	1800	1650	1760	3520	5600	2450	2680
2700	1680	3650	3240	5850	3150	1860	2425
4520	3275	4215	3760	1950	1850	3750	2825
4500	3800	4300	2750	4370	3350	2375	3215

6. Discuss the fundamental rules of code construction.
   Discuss with the help of suitable examples various steps involved in data processing.
8. Indicate the diagrams you would consider most appropriate to use for representing each of the following classes of statistical data stating briefly the reason for choice
   a) Distribution of a large number of candidates according to the number of marks scored by each at a public examination.
   b) Distribution by age, sex and civil condition of persons enumerated at the census in 1981.
   c) Distribution of assets of All Indian Insurance Companies put together as at Jan. 19, 1996.
   d) Total value of India Exports and Imports during the years 1995-1996.
9. Draw a pie chart for comparing the various costs (rs. in lakhs) of the house building activities in two periods, 1984 and 1989.

	1984	1989
Land cost	0.75	.00
Material cost	1.00	.75
Labour cost	0.60	.00
Fixtures & furnitures	0.40	.75
Miscellaneous	0.25	.50

Objectives

After studying this unit, you should be able to:

- distinguish between descriptive and inferential statistic
- explain the usefulness of univariate tables
- interpret the results of descriptive measures related to univariate analysis
- discuss various inferential techniques as applicable to univariate analysis

Structure

10.1 Introduction
10.2 Distinction between Description and Inference
10.3 Framework for Analysis
10.4 Analysis of Summary Measures of Sample Data
10.5 Results from Inferential Statistics
10.6 Summary
10.7 Self-Assessment Questions
10.8 Further Readings

10.1 INTRODUCTION

The unit on data processing covered coding, tabulation and presentation of data collected from both primary and secondary sources. Once the data has been collected, it needs to be analysed for making sound marketing decisions. Analysis can be done on the basis of one variable, two variables & more than two variables at a time. When we are analysing one variable at a time, the analysis is named as *Univariate Analysis*. When the association between two variables is of concern, the type of analysis is termed as *Bivariate Analysis*. However, when simultaneous relationship between three or more variables is of concern to the researcher, a *Multivariate Analysis* is taken. In this unit we will be concerned with description and inference from the sample data considering one variable at a time. The association between two variables will be covered in Unit 11, whereas blocks 4 & 5 covering multivariate analysis will deal with three units each describing the simultaneous relation between variables.

Activty 1

Define following concepts withthe help of examples

1. Univariate Analysis

.....

2. Bivariate Analysis

.....

3. Multivariate Analysis

....

10.2 DISTINCTION BETWEEN DESCRIPTION & INFERENCE

Before we proceed further, it is essential to distinguish between descriptive and inferential statistic. The descriptive statistic deals with the summary measures relating to the sample data. The examples are as under:

1. *What is the average income of the sampled respondents?*
2. *What is the dispersion of ages of the sampled respondents?*
3. *What is the distribution of respondents by sex?*
4. *Which income group has maximum number of users of a particular brand of product?*
5. *What percentage of respondents are married?*
6. *Which newspaper is read by majority of the sampled respondents?*
7. *Is there any association between the income and willingness to buy of the product?*
8. *What proportion of consumers of a particular brand of product are professional?*
9. *Is there any relationship between age and the frequency of purchase of a particular brand of a product?*

The inferential statistic deals with drawing inferences about the population parameters on the basis of sample results. The concern, here, is to generalize the results obtained from sample. Here the analysis is based on probability theory and a necessary condition for the use of inferential statistic is the randomization of the sample. The following are the examples of some questions which are addressed under inferential statistic.

1. *Is the average income of population greater than Rs. 9,000 per month?*
2. *Does the difference in the frequency of purchase of a particular brand of a product vary with age significantly?*
3. *Does the income of users and non-users of the product vary significantly?*
4. *Is the decline in sales of the product significant?*

Both descriptive and inferential statistic are important in making sound marketing decisions. The researcher must be aware of the type of analysis to be carried out since both descriptive and inferential techniques vary with the levels of measurement. Appendix 1 and 2 of this unit review briefly such techniques for carrying out univariate analysis.

Activity 2

Distinguish between descriptive and inferential analysis of data.

.....

.....

.....

.....

10.3 FRAMEWORK FOR ANALYSIS

Once the raw data is collected it needs to be brought into array and presented in the form of tables and graphs so that further analysis could be carried out with ease. Having done this, one would be interested in finding out the summary measures of sample data viz. measures of central tendency, dispersion etc. The marketing researcher would also like to use inferential techniques to draw inferences about population on the basis of sample results to arrive at generalization. As mentioned in Unit 7 on Attitude Measurement & Scaling, the method of analysis depends upon the levels of measurements used in measuring the data. It was mentioned that there are four types of measurements namely Nominal Scale, Ordinal Scale, Interval Scale & Ratio Scale. The statistical methods applicable under each of these scales of measurement have already been summarized in Table 1 of Unit 7.

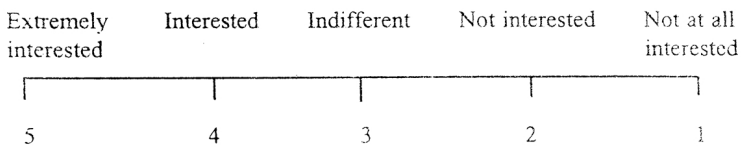
We will take up an example to illustrate the use of descriptive and inferential analysis for our purpose.

Example: XYZ company is considering the introduction of a new packaging design for its product. To date, the product was packed in a glass jar. The company wishes to replace it

with plastic containers. However, before doing so it wants to ascertain the opinion from its customers regarding their preference for new packaging or the inclination to stick to the old one.

The product of the company is sold in all parts of the country. It decides to take a random sample of 60 covering customers from northern, southern, eastern and western part of the country. The question asked to the customers is

How interested are you in buying the product if the packaging for the product is switched to the plastic container?



The other variables on which data collected from respondents included sex, marital status age, education and monthly income

The collected data is presented below in Table 1

Table 1: : SURVEY DATA ON PREFERENCE FOR NEW PACKAGING DESIGN

RESPONDENT NUMBER	PREFERENCE	PREFERENCE CODE	SEX	MARITAL STATUS	AGE IN YEARS	EDUCATION CO	ACTUAL INCOME (RS.)	INCOME CODE	REGION CODE
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	3	Not interested	M	M	25	11	2800	21	31
2	2	Not interested	M	M	27	13	3782	22	33
3	2	Not interested	F	S	28	13	6072	23	32
4	5	Interested	M	S	24	14	9050	24	34
5	3	Not interested	M	M	30	12	4982	22	32
6	1	Not interested	M	S	35	12	5375	23	31
7	2	Not interested	F	M	39	12	4210	22	34
8	4	Interested	F	S	37	14	7563	23	31
9	4	Interested	M	M	36	13	5384	23	33
10	2	Not interested	F	M	29	13	3218	22	32
11	2	Not interested	M	S	41	11	1976	21	31
12	1	Not interested	M	S	43	11	3723	22	32
13	2	Not interested	M	M	40	12	2187	21	34

14	5	Interested	F	M	31	13	9572	24	34
15	4	Interested	M	S	35	13	5689	23	33
16	3	Not interested	F	S	45	12	5791	23	32
17	2	Not interested	F	M	46	11	2500	21	31
18	4	Interested	M	S	38	13	3780	22	33
19	5	Interested	F	M	40	14	5004	23	34
20	4	Interested	M	S	39	14	87	24	32
21	2	Not interested	M	S	27	13	4611	22	31
22	3	Not interested	M	M	27	13	6666	23	34
23	4	Interested	F	S	31	13	6129	23	33
24	5	Interested	M	M	24	14	8289	24	32
25	2	Not interested	F	S	32	13	7270	23	33
26	4	Interested	M	M	38	14	6235	23	33
27	2	Not interested	F	M	25	13	4219	22	31
28	3	Not interested	F	M	29	12	4136	22	32
29	4	Interested	M	S	40	13	3784	22	34
30	3	Not interested	F	S	45	14	10000	24	33
31	4	Interested	F	M	35	13	6782	23	32
32	5	Interested	F	S	32	14	6115	23	31
33	1	Not interested	M	S	29	12	7068	23	32
34	2	Not interested	M	M	42	12	2991	21	34
35	1	Not interested	F	M	27	11	3787	22	33
36	3	Not interested	M	M	29	11	2736	21	32
37	4	Interested	M	S	28	11	5436	23	34
38	3	Not interested	M	M	41	12	1985	21	31
39	4	Interested	M	S	43	13	4803	22	33

40	3	Not interested	F	S	50	12	1999	21	31
41	5	Interested	M	M	52	13	5072	23	32
42	3	Not interested	F	M	47	14	5815	23	33
43	4	Interested	F	S	30	13	3925	22	34
44	2	Not interested	M	M	53	11	2673	21	32
45	1	Not interested	M	S	39	12	4271	22	31
46	4	Interested	F	M	55	13	3791	22	34
47	3	Not interested	M	S	49	11	4213	22	33
48	4	Interested	F	M	38	12	5824	23	32
49	3	Not interested	M	S	27	13	3270	22	34
50	4	Interested	F	S	46	13	6184	23	31
51	3	Not interested	M	S	47	14	4634	22	31
52	2	Not interested	F	M	38	14	6224	23	33
53	1	Not interested	F	M	28	13	3182	22	32
54	5	Interested	M	M	43	11	8467	24	31
55	3	Not interested	F	S	39	12	2789	21	34
56	4	Interested	M	S	44	12	6972	23	31
57	5	Interested	F	M	29	13	8131	24	33
58	4	Interested	M	S	41	11	2835	21	32
59	5	Interested	F	S	26	12	5138	23	34
60	4	Interested	M	M	32	13	9220	24	33

From the above table, the following points may be noted:

(i) Column one indicates the respondent number.

(ii) Column two indicates the rating given by the respondents for their preference or dislike for the new packaging on a 5-point scale. As discussed earlier the scale value of 5 indicates that the respondent is extremely interested, whereas the scale value of 1 represents that respondent is not at all interested in the new packaging.

(iii) In column three, the respondents are divided into two categories - Interested & Not interested. All those respondents whose score were 4 and 5 are grouped under "Interested" and those whose score were 1, 2, 3 are grouped as "Not-interested".

(iv) Column four indicates the sex of the respondent, 'M' stands for male whereas 'F' stands for Female respondent.

(v) Column five exhibits the marital status of the respondent. If a respondent is married he/she is coded as 'M', whereas if the respondent is single he/she is coded as 'S'

(vi) Column six gives the age of the respondent measured as a ratio scale variable.

(vii) Column seven indicates the education category to which the respondent belongs. There are four categories for this variable namely below higher-secondary, higher secondary, graduation and post graduation. The code for below higher secondary is 11, for higher secondary 12, for graduation 13 and for post graduation 14. This is an ordinal scale data and therefore may also be used as a nominal variable.

(viii) Column eight gives the monthly income of the respondent. Again this is a ratio scale data.

(ix) In column nine, the data on monthly income as given in column eight has been classified into four categories viz. less than Rs. 3000 coded as 21, Rs. 3000 - Rs. 4999 coded as 22, Rs. 5000 - Rs. 7999 coded as 23 and Rs. 8000 and above coded as 24. Again this is the case of ordinal scale data which can also be converted into nominal scale data.

(x) Column ten indicates the region from where the respondent is selected. Four regions namely, north, south, east and west were considered with respective codes given by 31, 32, 33 and 34.

10.4 ANALYSIS OF SUMMARY MEASURES OF SAMPLE DATA

The concern is to analyse sample data with measures of central tendency and dispersion.

1. **Preference:** The preference data as given in column two is an interval scale data and therefore the measures of central tendency and dispersion applicable in this case are mean and standard deviation. Of course, all other measures applicable in the case of ordinal and nominal scale data can also be used in this case. The summary results corresponding to preference data (column 2) is given below in Table 2

Table 2 : SUMMARY RESULTS OF PREFERENCE DATA

Measure	Value	Measure	Value
Mean	3.183	Std. Deviation	1.228
Median	3.000	Skewness	-.192

The above table indicates a mean value of 3.183 & median value of 3.0. The standard deviation & skewness are 1.228 & -.192 respectively. The results indicates that the distribution is skewed to the left, has a mean value of 3.183 which is slightly above "indifferent" and is on the side of "interested" category. Median falls in the neutral category. Mode should not be applied to ordinal or interval data unless these have been grouped first.

The same preference data when categorized as "interested" and "not interested" as given in column number three gives the following results. Out of 60 respondents, 27 prefer the new packaging design whereas 33 are either neutral or do not want the

change in design. This being nominal scale data, we can only compute mode of the distribution which falls in the category of "not interested". The measure of dispersion applicable in this case is relative frequencies which indicate that 55% of the respondents are not interested in the new design.

2. Sex: The Table 3 below gives the distribution of respondent with respect to sex

Table 3 : FREQUENCY DISTRIBUTION OF RESPONDENT BY SEX

Sex	Frequency	Per cent
Female (F)	27	45.0
Male (M)	33	55.0
Total	**60**	**100.0**

The above table indicates that there are 33 male respondents comprising 55% of the sample, whereas there are only 27 female respondents. This being a nominally scaled data, the mode of the distribution corresponds to "male" category.

3. Marital Status: The Table 4 below gives the marital status of the sample respondents.

Table 4 : FREQUENCY DISTRIBUTION OF RESPONDENTS BY MARITAL STATUS

Marital Status	Frequency	Per cent
Married (M)	30	50.0
Single (S)	30	50.0
Total	**60**	**100.0**

The above table indicates that there is an equal number of married and single respondents.

4. Age: The Table 5 below gives the results corresponding to age of the respondents.

Table 5 : SUMMARY RESULTS CORRESPONDING TO AGE VARIABLE

Measure	Value	Measure	Value
Mean	36.417	Std. Deviation	8.164
Median	37.500	Skewness	0.309

The results indicates that the average age of the respondent is 36.417 years. The median value of 37.5 indicates that 30 respondents are below the age of 37.5 whereas the remaining 30 are above the age of 37.5. The distribution is positively skewed.

5. **Education:** The data on education measured in nominal scale is given in Table 6 below:

Table 6 : FREQUENCY DISTRIBUTION OF RESPONDENTS BY EDUCATION

Education	Frequency	Per cent	Cum Per cent
Below higher secondary	11	18.3	18.3
Higher secondary	15	25.0	43.0
Graduation	23	38.0	81.0
Post-graduation	11	18.3	100.0
TOTAL	**60**	**100.0**	

It may be observed that about 57% of the respondent have an education level of Graduation and above. The mode of the distribution is the category "Graduation". Here we are treating this as a nominally scaled data. It can however be considered as ordinal scale data under the assumption that for doing post- graduation one has to spend more number of years than for a graduation and so on.

6. **Income:** Income data expressed in ratio scale is given in column 8 of Table 1. The summary measures computed for this is given in the following Table 7.

Table 7 : SUMMARY MEASURES FOR INCOME VARIABLE

Measure	Value	Measure	Value
Mean	5150.983	Std. Deviation	2081.689
Median	4993.000	Skewness	0.500

The mean income is 5150.983, the median being Rs. 4993 expressing that 50% of the respondent have income below it and 50% have income above it. The distribution is skewed to the right. The results for income variable when expressed in coded form using coding scheme already explained is given below in Table 8.

Table 8 : FREQUENCY DISTRIBUTION OF RESPONDENTS BY INCOME

Value	Frequency	Per cent	Cum Per cent
21	11	18.3	18.3
22	19	31.7	50.0
23	22	36.7	86.7
24	8	13.3	100.0
TOTAL	**60**	**100.0**	

It can be noted that the mode in the above table is the income category Rs. 5000 - 7999. The data may be treated as ordinal scale data and the median of the distribution lies in the income group Rs. 3000 - Rs. 4999.

7. **Region:** The regionwise distribution of respondents is given in Table 9 below:

Table 9: FREQUENCY DISTRIBUTION OF RESPONDENT BY REGION

Region	Frequency	Percent
North	15	25.0
South	16	26.7
East	15	25.0
West	14	23.3
Total	**60**	**100.0**

It may be noted that the maximum number of respondents are from south whereas the minimum number is from west. However, there does not seem to be much variation among the number of respondents from the four regions.

The analysis carried out above by forming one-way table is useful for the following purposes:

1. These tables provide valuable insights through comparison with other relevant distributions. For example our sample results as given in Table 3 indicates that there were 55% male and 45% female respondents. However, if this distribution is not in tune with the population distribution, we may infer that the sample is not representative of the population.

2. One way table can also suggest useful variable transformation. Consider, e.g. the results indicated in Table 7 for the income variable. It is seen that the distribution is skewed. Therefore one needs to transfer the data into appropriate transformation so as to enable the analyst to carry out appropriate statistical tests which are otherwise not possible in the case of skewed distribution.

Activity 3

How will you use the results of univariate anslysis to judge the represeemtauves of a sanple

.....
.....
.....
.....

Activity 4

How are (1) measures of central tendency and (2) measures of dispersion both neccssory to describe a sample?

.....
.....
.....
.....

10.5 RESULTS FROM INFERENTIAL STATISTIC

As mentioned earlier, inferential statistic attempts to generalize the results from sample to population. We will try to answer following questions:

1. Is the distribution *of respondents across various regions uniform?*
2. *Are the respondents uniformly distributed across 4 education level categories?*
3. *Is the proportion of respondent preferring the new scheme different from .50?*
4. *Is the average income of all respondents significantly different from Rs. 4500?*

To answer first question we test the following hypothesis.

H_0 *(Null-hypothesis):* The distribution of respondents preferring the new packaging design is uniform across all the four regions.

H_1 *(Alternative-hypothesis):* The distribution of respondents across the four regions is not uniform.

We are given the observed number of respondents belonging to various regions in Table 10. If the null-hypothesis is true, there should be 15 respondents from each of the four regions. Further to test the hypothesis we need to carry out the test of goodness of fit by computing Chi-square statistic. The results are given below in Table 10.

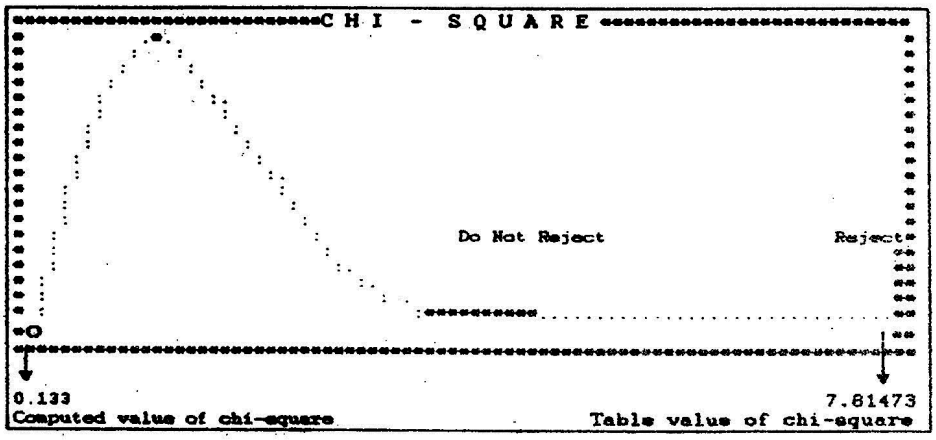
Table 10: TEST OF GOODNESS OF FIT FOR REGIONS

S.No.	Observed Frequency (Oi)	Expected Frequency (Ei)
1	15	15
2	16	15
3	15	15
4	14	15
Total	**60**	**60**

Computed chi-square: $\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ = 0.1333

Critical chi-square = 7.8147
(assuming level of significance = 0.05)

As computed chi-square value is less then the critical chi-square value, we do not reject null-hypothesis. This shows that the distribution of respondents across various region is uniform. The results may be shown diagramatically as follows:



To answer second question the following hypothesis needs to be tested.

H_0 (Null-hypothesis): The distribution of respondents preferring the new packaging design is uniform across all the four education categories.

H_1 (Alternative-hypothesis): The distribution of respondents across the four education categories is not uniform.

The hypothesis can be tested using chi-square test for goodness of fit. Since there are 60 respondents, therefore, under the assumption that the null-hypothesis is true, it may be expected that each of the categories has 15 respondents.

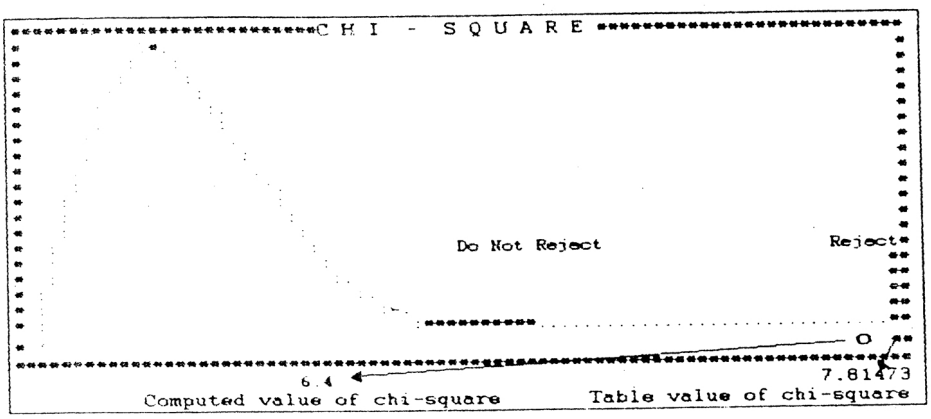
Table 11 : TEST OF GOODNESS OF FIT FOR EDUCATIONAL CATEGORIES

Education Level	Observed Frequency (Oi)	Expected Frequency (Ei)
Below higher secondary	11	15
Higher secondary	15	15
Graduation	23	15
Post graduation	11	15
Total	**60**	**60**

Computed chi-square: $\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ = 6.4000

Critical chi-square = 7.8147
(assuming level of significance = 0.05)

As computed chi-square is less than the critical value of chi square we do not reject the null-hypothesis. Therefore, this shows that the distribution of respondents across various educational categories is uniform. The results may also be shown diagrammatically as:



To answer the third question, we test the following hypothesis.

H_0 *(Null-hypothesis):* The proportion of respondents preferring the new packaging design equals .50. ($\pi = 0.50$)

H_1 *(Alternative-hypothesis):* The proportion is different from .50. ($\pi \neq 0.50$)

The sample proportion $p = \frac{27}{60} = .45.$

As sample is large (n = 60) and both np & n(1-p) are greater than 5, we may use normal approximation to binomial distribution and apply Z test.

Let us assume 5% level of significance.

Critical z value = 1.96.

Lower limit of acceptance region $= \pi_{Ho} - Z_{\alpha/2}\,\sigma_p$

$= .50 - 1.96\,(.0642)$

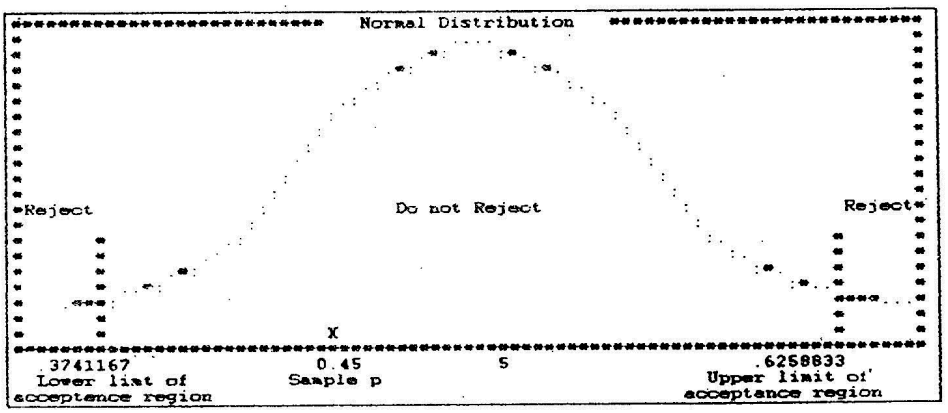
$= .3741$

Upper limit of acceptance region $= \pi_{Ho} + Z_{\alpha/2}\,\sigma_p$

$= .50 + 1.96\,(.0642)$

$= .6259$

Since p lies in the upper and lower limit of acceptance region, we do not reject H_0. This is shown graphically as follows.



The fourth question may be answered by testing the following hypothesis.

H_0 (Null-hypothesis): Average income of all the respondents equals Rs. 4500 ($\mu = 4500$).

H_1 (Alternative-hypothesis): The average income of all the respondents is different from Rs.4500 ($\mu \neq 4500$).

We use data given in column eight of Table 1 to test this hypothesis at 5% level of significance. The sample size being large (n = 60), we may use z test for the purpose.

Sample mean $\bar{x}$ = Rs. 5150.9834.

Standard error of mean ($\sigma_{\bar{x}}$) = 268.7448.

Lower limit of acceptance region = $\mu_{Ho} - Z_{\alpha/2}\,\sigma_{\bar{x}}$

= 4500 - 1.96 (268.7448)

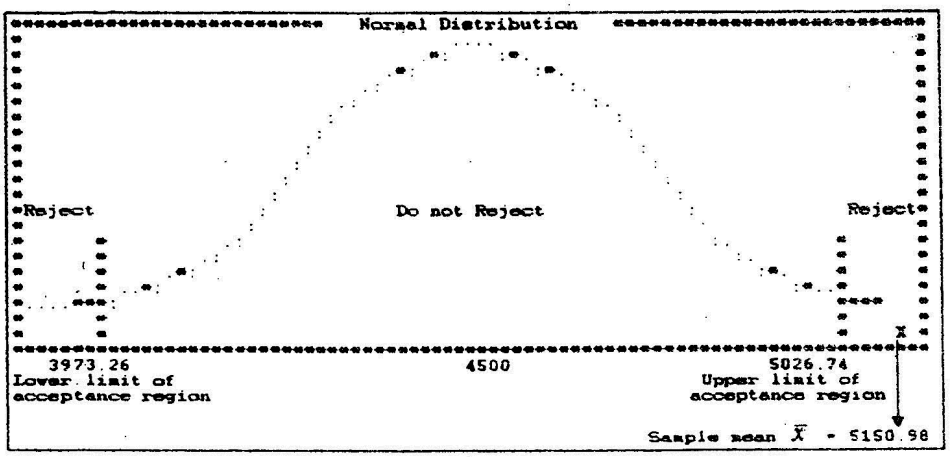
= 3973.26

Upper limit of acceptance region = $\mu_{Ho} + Z_{\alpha/2}\,\sigma_{\bar{x}}$

= 4500 + 1.96 (268.7448)

= 5026.74

Since $\bar{x}$ lies in the acceptance region we accept the null-hypothesis. This is shown graphically as follows:



Activity 5

From pust records it is known that the average life of a battery used in a diginal clok iss sd305 day. The lices of the batteries are normally distributed. The baucry was recenty inodificd to last longer . a sample of the modified batteries was tested lt was discovered that the mean life was 311 days and the sample standard deviation was 12 days. At the 0.05 level of significancc did the modification increase the mean life of the battery

.....
.....
.....
.....

Activity 6

A company markets car tyres. Their lives are normally distributed with a mean of 40,000 kilometres and standard deviation of 3,000 kilometers. A change in the production process is believed to result in a better product. A test sample of 64 new tyres has a mean life of 41,200 km. Can you conclude that there is no significant difference between the new product and current one. You may use a 5% level of significance.

.....
.....
.....
.....

Activity 7

A high-speed automatic machine mass produces a small washer. Past experience reveals that 70 per cent of each day's production is perfect. Most of the remaining washers have a rough burr, which must be filed off before the washers can be inserted in the assembly. In an attempt to increase the per cent of production that is perfect, the machine was modified somewhat. A sample of 100 washers was then checked, and it was found that 72 percent were perfect. The boss thinks that there has been no change. The plant manager, however, believes that the production of the modified machine has definitely improved product quality.

Description & Laference From Samole Data

Category Label	Code	Absolute Frequency	Relative Frequency (Per cent)	Cumulative Frequency (Per cent)
Strongly disagree	1	110	24.44	24.44
Neutral	2	90	20.00	44.44
Strongly agree	3	250	55.56	100.00
Total		450		

4. The manager of a departmental store hypothesized the number of customers entering the store alone is 1.5 times the number entering in groups. A sample of 240 customers yielded the following results.

Entering Alone	140
Entering in Group	100

Is the manager's hypothesis about population of custamers correct

5. What is the basic use of a chi-square goodness of fit test? How is the value of the test statistic calculated? How are the expected frequencies determined?

6. What is the appropriate test statistic for making inferences about a population mean when the variance is known? When the population variance is unknown? Suppose the population variance is unknown, but the sample size is large, what is the appropriate procedure then?

7. The number of demands for a particular spare part in a shop was found to vary from day to day. In a sample study the following information was obtained:

Days	:	Mon	Tue	Wed	Thu	Fri	Sat
No. of parts demanded	:	124	125	110	120	126	115

Test the hypothesis that the number of parts demanded doesn't depend upon the day of the week. You may use 5% level of significance.

8. The records of the owner of a fleet of trucks revealed that the average life of a set of spark plugs is 22,100 miles. The distribution of the life of the plugs is near normally distributed. A spark plug manufacturer claimed that its plugs has an average life in excess of 22,100 miles. The fleet owner purchased a large number of sets. A sample of the life of 18 sets revealed that the sample average life was 23,400 miles; the sample standard deviation was 1,500 miles. Is there enough evidence to accept the manufacturer's claim at the 0.05 level?

9. A food processor is concerned that the 16-ounce can of sliced pineapple is being overfilled. The quality control department took a random sample of 50 cans and found that the arithmetic mean weight was 16.05 ounces, with a sample standard deviation of 0.03 ounces. At the 5 per cent level of significance, can the hypothesis that the mean weight is equal to 16 ounces be rejected?

10. In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers? You may use 5% level of significance.

10.8 FURTHER READINGS

Beri, G. C. *"Marketing Research — Text and Cases"* Tata McGraw-Hill Publishing Co. Ltd. (1st Edition).

Green, Paul E. and Donald S. Tull *"Research for Marketing Decisions"* Prentice-Hall of India Pvt. Ltd. (4th Edition).

Kinnear, Thomas C. and James R. Taylog, *"Marketing Research — An Applied Approach"* McGraw-Hill International Editions (3rd Edition).

Luck, David J. and Ronald S. Rubin, *"Marketing Research"* Prentice-Hall of India Pvt. Ltd. (7th Edition).

Majumdar, Ramanuj *"Marketing Research — Text, Applications and Case Studies"* Wiley Eastern Ltd. (1st Edition).

Westfall, Boyd and Stasch. *"Marketing Research — Text and Cases"* Richard D. Irwin, Inc. (6th Edition).

APPENDIX - 1

This appendix gives a brief review of measures of central tendency and dispersion. They have already been discussed in detail in Unit numbering 7 & 8 of the course on Quantitative Analysis for Managerial Applications (MS-8). You are therefore advised to go through it in case you have forgotten it.

Measures Of Central Tendency:

The principal measures of central tendency are arithmetic mean, the median and the mode.

- **Arithmetic Mean:** This measure of central tendency is used in case of interval or ratio scale data. In case of grouped data, the arithmetic mean may be obtained as:

- $$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{n}$$

Where $\overline{X}$ = the sample mean
f_i = frequency of the ith class
X_i = Midpoint of the ith class
k = Number of classes
n = Total number of observations in the sample.

The population mean (μ) may be computed as:

$$\mu = \frac{\sum f_i X_i}{N}$$

Where μ = The mean of population
N = Number of observations in the population

Median: It is a measure of central tendency of a distribution. It is that value in the distribution such that 50 per cent of the observations are below it and 50 per cent are above it. Assuming odd number of observations in the distribution, it is the middle value of the distribution when the data is arranged in the ascending order of magnitude. When the number of observations are even, median is the mean of two middle values. In case of grouped data, the median is obtained by using the following formula :

$$M = l_i + \frac{\left(\frac{n}{2}\right) - CF}{f} \times (i)$$

Where M = Median

l_i = Lower limit of the median class

n = Total number of frequencies in the distribution

CF = Cumulative frequencies for the class immediately below the class containing the median.

i = Size of the interval of the median

f = Frequency of median class

Mode: It is usually determined when data is categorised into frequency distribution. Mode represents the category having the highest number of observations. The mode may be computed from the grouped data as follows:

$$M_o = l_i + \frac{d_1}{d_1 + d_2}(i)$$

Where M_o = Mode

l_1 = Lower limit of the modal class.

d_1 = Difference between the frequency of the modal class and the frequency of the class immediately preceding it in the distribution.

d_2 = Difference between the frequency of the modal class and frequency of the class immediately following it in the distribution.

I = size of the interval of the modal class.

Measures of Dispersion: Measures of central tendency alone are not sufficient as they give a very inadequate description of the sample data. This is because these measures are not concerned with the spread of the distribution. To measure spread or dispersion of the distribution some other measures like range, mean deviation, standard deviation, variance and co-efficient of variation are used.

- **Range:** This is the simplest measure of dispersion used for interval or ratio scaled data. It is measured as the difference between the highest and the lowest value in the distribution.

- **Variance and Standard Deviation:** The variance may be defined as the average of squared deviations of the observation values from the mean of the distribution. Standard deviation is the positive square root of the variance. The population standard deviation (σ) may be computed as:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}$$

Where X_i = the value of ith observation
μ = population mean
N = Number of observations in the population.

For computing standard deviation from the sampled data, following formula may be used.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

Where $\overline{X}$ = Sample mean
n = Number of observations in the sample.

Co-efficient of Variation: It is a relative measure of dispersion to compare the variability of more than one distribution. This is computed for ratio-scaled data. The formula for its computation is as given below:

$$CV = \frac{\sigma}{\mu}(100) \quad or \quad \frac{s}{\overline{X}}(100)$$

Where CV= Coefficient of variation
σ = Standard deviation of the population.
μ = Mean of the population
s = Standard deviation of the sample.
$\overline{X}$ = Mean of the sample.

To compare variability of various distribution, the coefficient of variation corresponding to each distribution is computed and the one having the minimum value is considered to be the more homogeneous.

APPENDIX - 7

This appendix gives a brief review of the inferential techniques for inferring the parameters of a single population (for details refer Unit 15 of MS-8 Course).

1. **Test Of Mean (Single population):** On the basis of sample drawn from the population one needs to infer about the population parameter. The parameter of interest could be population mean. One needs to carry out an appropriate statistical test of significance for testing the hypothesis concerning population mean. One can consider two cases — sample size being large ($n > 30$) and small sample ($n \leq 30$). If n is large the sampling distribution of sample means follows a normal distribution as per central limit theorem. Therefore one can go for Z test. However when sample size is small, assuming that the samples are drawn from a normal population one may have to choose between Z and t test depending upon whether population standard deviation (σ) is known or not. In case σ is known one should go for Z test otherwise t test is carried out.

Let us assume that we want to test the hypothesis that the population mean (μ) takes a specified value against an appropriate alternative hypothesis. For the sake of simplicity let us assume a two tailed alternative. Therefore the null and alternative hypothesis can be written as given below.

H_0 (Null hypothesis)	μ	=	μ_0
H_1 (Alternative hypothesis)	μ	≠	μ_0

Let us assume a large sample and the level of significance being equal to α. The test procedure involves the following steps:

1. Finding the sample mean $\bar{x}$

2. Computing the standard error of mean $\left(\sigma_{\bar{x}} \quad \frac{\sigma}{\sqrt{n}}\right)$

3. Finding the value of $Z_{\alpha/2}$ from the standard normal table.

4. Finding the upper and lower limit of the acceptance region as given below:

   Upper limit of acceptance region = $\mu_{H_0} + Z_{\alpha/2} \; \sigma_{\bar{x}}$

   Lower limit of acceptance region = $\mu_{H_0} - Z_{\alpha/2} \; \sigma_{\bar{x}}$

   where μ_{H_0} represents the value of μ under the assumption that the null hypothesis is true.

   Examining whether sample mean $\bar{x}$ lies in the upper and lower limit of the acceptance region. If it does, we accept the null hypothesis otherwise reject it.

In case of small sample drawn from normal population with unknown standard deviation S the same hypothesis can be tested with the help of t statistic. The procedure is outlined below.

1. Find the sample mean $\bar{x}$

2. Find an estimate of population variance. The sample variance can be used to estimate population variance. The formula for computing sample variance is given as

$$s^2 = \frac{1}{n-1}\sum(Xi - \overline{X})^2$$

   s^2 is used as an unbiased estimator of σ^2.

3. The standard error of mean is estimated as $\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}$ where ^ stands for estimates.

4. Find the value of $t_{\alpha/2}$ with n - 1 degrees of freedom from t distribution table.

5. Find the upper and lower limit of the acceptance region as given below:

Upper limit of acceptance region $= \mu_{H_0} + t_{\alpha/2} \hat{\sigma}_{\bar{x}}$

Lower limit of acceptance region $= \mu_{H_0} - t_{\alpha/2} \hat{\sigma}_{\bar{x}}$

where μ_{H_0} represents the value of μ under the assumption that the null hypothesis is true.

Examine whether sample mean $\bar{x}$ lies in the upper and lower limit of the acceptance region. If it does, we accept the null hypothesis otherwise reject it.

In case of one tail test, the procedure remains same except that one computes either upper or lower limit of the acceptance region depending upon whether alternative hypothesis is given by $H_1 : \mu > \mu_0$ or $H_1 : \mu < \mu_0$. For example, in case of Z test the upper limit will be computed as

$$\mu_{H_0} + Z_\alpha \sigma_{\bar{x}}$$

2. **Test of Proportion:** In order to test the hypothesis that population proportion (P) takes a specified value P_0 against a two sided alternative, we test the following hypothesis.

H_0 (Null Hypothesis) : $\pi = \pi_0$
H_1 (Alternative Hypothesis) : $\pi \neq \pi_0$

Assuming large sample such that both NP & nq > 5 (q = 1-p), we may use normal approximation to binomial distribution to test the above hypothesis. It can be shown that

1. $E(P) = \pi$

2. $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

where p denotes sample proportion and σ_p denotes standard error of sample proportion.

To test the null hypothesis we compute upper and lower limit of acceptance region as given below

Upper limit of acceptance region = $\pi_{H_0} + Z_{\alpha/2} \sigma_p$

Lower limit of acceptance region = $\pi_{H_0} - Z_{\alpha/2} \sigma_p$

If sample proportion *p* lies in the lower and upper limit of the acceptance region we accept H_0 otherwise reject.

For a one tail test the procedure is similar to as explained in the case of tests of mean.

Test of Goodness of Fit: We will explain how chi-square statistic can be used to test whether samples are drawn from a uniform population. One can compute the expected frequencies under the assumption that the null hypothesis is true. The hypothesis may be stated as:

H_0 (Null Hypothesis):	The sample is drawn from a population which follows a normal distribution.
H_1 (Alternate Hypothesis):	The data does not belong to the uniform distribution.

One can compute chi-square statistic by using the following formula:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Here chi-square has k-1 degrees of freedom since we are assuming that the data is categorized into k groups. For a given level of significance one compares the table value of chi square with k-1 degrees of freedom to the computed value of chi-square and if tabulated chi-square is less than the computed chi-square we accept the null-hypothesis of uniformity of distribution.

UNIT 11 ANALYSIS OF ASSOCIATION

Objectives

After studying this unit, you should be able to:

- define a contingency table
- compaers cross-tabulation tables
- Descuss the appllcation of chi-square test to examine the independence of two nomically scaled variables
- explain different measures to ascertain the strength of relationship between two nominal variables
- carry out appropriate tests of difrerenee between two population parameters
- distinguish between correlation and regression and their use in measuring association between variables.

Structure

11.1 Introduction

11.2 Cross-tabulation

11.3 Chi-square Test for Analysis of Association

11.4 Strength of Assocaition between two Nominal Variables

11.5 Correlation Coefficient

11.6 Simple Linear Regression

11.7 Analysis of Difference

11.8 Summary

11.9 Self-Assessment Questions

11.10 Further Readings

11.1 INTRODUCTION

We discussed the description and inference from statistical data by taking one variable at a time in unit number 10 Our concern in this unit is on bivariate analysis of data where we take the simultaneous relationship between two variables. We are concerned with the cross-tabulation of data and its interpretation, analysis of association by using chi-square test and correlation coefficient. The cause and effect relationship between two variables will be examined by using simple linear regression model. We would carry out tests of statistical significance on the samples drawn from bivariate population viz. tests for the differences of means & differences of proportions. The use of various techniques will be illustrated with the help of survey data.

11.2 CROSS TABULATION

A matrix display of the categories of two nominal scaled variables, containing frequency counts of number of subjects in each bivariate category is called cross-tabulation table or contingency table. For preparing cross tables to examine simultaneous relations between two variables, we will consider the data given in Table 1 of the last unit. For the sake of convenience, the data is reproduced below.

TABLE 1 : SURVEY DATA ON PREFERENCE FOR NEW PACKAGING DESIGN

RESPONDENT NUMBER	PREFERENCE	PREFERENCE CODE	SEX	MARRITAL STATUS	AGE IN YEARS	EDUCATION CODE	ACTUAL INCOME (RS.)	INCOME CODE	REGION CODE
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	3	Not interested	M	M	25	11	2800	21	31
2	2	Not interested	M	M	27	13	3782	22	33
3	2	Not interested	F	S	28	13	6072	23	32
4	5	Interested	M	S	24	14	9050	24	34
5	3	Not interested	M	M	30	12	4982	22	32
6	1	Not interested	M	S	35	12	5375	23	31
7	2	Not interested	F	M	39	12	4210	22	34
8	4	Interested	F	S	37	14	7563	23	31
9	4	Interested	M	M	36	13	5384	23	33
10	2	Not interested	F	M	29	13	3218	22	32
11	2	Not interested	M	S	41	11	1976	21	31
12	1	Not interested	M	S	43	11	3723	22	32
13	2	Not interested	M	M	40	12	2187	21	34
14	5	Interested	F	M	31	13	9572	24	34
15	4	Interested	M	S	35	13	5689	23	33
16	3	Not interested	F	S	45	12	5791	23	32
17	2	Not interested	F	M	46	11	2500	21	31
18	4	Interested	M	S	38	13	3780	22	33
19	5	Interested	F	M	40	14	5004	23	34
20	4	Interested	M	S	39	14	8730	24	32
21	2	Not interested	M	S	27	13	4611	22	31
22	3	Not interested	M	M	27	13	6666	23	34
23	4	Interested	F	S	31	13	6129	23	33
24	5	Interested	M	M	24	14	8289	24	32
25	2	Not interested	F	S	32	13	7270	23	33
26	4	Interested	M	M	38	14	6235	23	33
27	2	Not interested	F	M	25	13	4219	22	31
28	3	Not interested	F	M	29	12	4136	22	32
29	4	Interested	M	S	40	13	3784	22	34

30	3	Not interested	F	S	45	14	10000	24	33
31	4	Interested	F	M	35	13	6782	23	32
32	5	Interested	F	S	32	14	6115	23	31
33	1	Not interested	M	S	29	12	7068	23	32
34	2	Not interested	M	M	42	12	2991	21	34
35	1	Not interested	F	M	27	11	3787	22	33
36	3	Not interested	M	M	29	11	2736	21	32
37	4	Interested	M	S	28	11	5436	23	34
38	3	Not interested	M	M	41	12	1985	21	31
39	4	Interested	M	S	43	13	4803	22	33
40	3	Not interested	F	S	50	12	1999	21	31
41	5	Interested	M	M	52	13	5072	23	32
42	3	Not interested	F	M	47	14	5815	23	33
43	4	Interested	F	S	30	13	3925	22	34
44	2	Not interested	M	M	53	11	2673	21	32
45	1	Not interested	M	S	39	12	4271	22	31
46	4	Interested	F	M	55	13	3791	22	34
47	3	Not interested	M	S	49	11	4213	22	33
48	4	Interested	F	M	38	12	5824	23	32
49	3	Not interested	M	S	27	13	3270	22	34
50	4	Interested	F	S	46	13	6184	23	31
51	3	Not interested	M	S	47	14	4634	22	31
52	2	Not interested	F	M	38	14	6224	23	33
53	1	Not interested	F	M	28	13	3182	22	32
54	5	Interested	M	M	43	11	8467	24	31
55	3	Not interested	F	S	39	12	2789	21	34
56	4	Interested	M	S	44	12	6972	23	31
57	5	Interested	F	M	29	13	8131	24	33
58	4	Interested	M	S	41	11	2835	21	32
59	5	Interested	F	S	26	12	5138	23	34
60	4	Interested	M	M	32	13	9220	24	33

Please note that the codes used in the table have the same meaning as explained in unit 10.

To interpret the results of cross tabulated data we will, as an illustration, try to answer the following questions:

1. *Divide the sample into two groups: (a) those showing interest in the new design and (b) those who are either indifferent or not interested in the new design. Cross-tabulate these two groups along-with education*

   (I) higher education - graduation and above and
   (II) lower education - below graduation

2. *Perform similar exercise to ascertain association between preference for new design and income level of respondents, taking the first income level in the poor class; second and third levels in the middle class; and the fourth one in the upper class.*

3. Perform sinilar exercise to asccrtuin associion between geographical locotion and prefeyence for new design as categorized in question

4. Prepanre a crass-tabulation for preference for nrw design versus age groups. wjen there are tow age groups- older respondents (40) ycars and above and vonmer respondems below 40 yrars

5. Prepare a criss-table for sex variable against preference for new .design

6. Also prepare cross-table for marital status variable against .Prefcrence for nrw design

Interpret all the above cross tabulations

Before we attempt to answer the above questions. it would be of help to identify dependent and independent variable. In the questions listed above the dependent variable is preference - indicated by interest shown by consumers in the new design. The independent variable varies from question to question. For example, it is education, income, geographical location, age, sex and marital status for question numbers 1, 2, 3, 4, 5 & 6 respectively.

Preference vs. Education: As indicated before, the sample is divided into two groups (a) those interested in the new design and (b) those who are either indifferent or not interested in the new design. This variable is cross tabulated alongwith education variable which is divided into two groups

(i) higher education - graduation and above &

(ii) lower education - below graduation.

The results of the cross-tabulation are presented in Table 2 below:

Table 2: CROSS TABULATION OF PREFERENCE WITH EDUCATION

Education Preference	Under-graduation	Graduation & Above	Row Total
Not Interested	20 (79.6%)	13 (38.2%)	33
Interested	6 (20.4%)	21 (61.8%)	27
Column Total	26 (100%)	34 (100%)	60

To interpret the results of the above table, a question which arises is - should we compute the percentages column wise or row wise? There is a general rule for computing percentages. ***The rule is that we should cast the percentages in the direction of causal variable.*** In Table 2 the causal variable is education whereas the dependent variable is preference. There are 26 consumers who are under-graduate and 20 of them are not interested in the new design. This means 76.9% of the undergraduates are not interested in the new design. We observe that there are 34 consumers whose qualification is graduation or above and 13 of them are not interested in the new design which accounts for 38.2% of the consumers with education - graduation and above. The results clearly indicates with the increase in education, the interest for the new design increases. Therefore we must cast the percentages column wise as shown in Table 2 above.

Preferences vs. Income: The income data is divided into three categories by taking the first level in the poor class; second and third level in the middle class; and the fourth level in the upper class. The preference data is categorized as already explained. The cross tabulated data for these variables is given in Table 3 below:

Table 3 : CROSS TABULATION OF PREFERENCE WITH INCOME

Preference / Income	Not Interested	Interested	Row Total
Poor Class	10 (99.01%)	1 (.09%)	11 (100%)
Middle Class	22 (53.66%)	19 (46.34%)	41 (100%)
Upper Class	1 (12.5%)	7 (87.5%)	8 (100%)
Column Total	33	27	60

There are 11 consumers in the income group categorized as poor class and out of which only 0.09% prefer the new design. In the middle income class, there are 41 consumers and 46.34% of them prefer the new design. There are 8 consumers in the upper income class and 87.5% of them have shown preference for the new design. The results clearly indicate that with increase in income, preference for the new design increases.

Preference vs. Geographical Location : In Table 1 we have data of consumers from northern, southern, eastern and western part of the country. When this data was cross-tabulated against the preference variable the results as indicated in Table 4 appears.

Table 4 : CROSS TABULATION OF PREFERENCE WITH GEOGRAPHICAL LOCATION

Preference / Region	Not Interested	Interested	Row Total
North	10 (66.67%)	5 (33.33%)	15 (100%)
South	10 (62.5%)	6 (37.5%)	16 (100%)
East	7 (46.67%)	8 (53.33%)	15 (100%)
West	6 (42.86%)	8 (57.14%)	14 (100%)
Column Total	33	27	60

There are 15, 16, 15 and 14 consumers from northern, southern, eastern and western regions respectively. There are 66.67% of the consumers from northern region, 62.5% from southern region, 46.67% from eastern region and 42.86% from western region are not interested in the new design. This shows that a higher proportion of customer from eastern and western part of the country prefer the new design than from northern and southern part of the country. However the difference in the preference does not seem to

be significantly very large. To test for the statistical significance of the association between these two variables, one needs to carry out chi-square test which will be explained in the next section.

Preference vs. Age Group: The data on age of the respondents given in Table 1 is divided into two groups - older respondents (40 years and above) and younger respondents (below 40 years). This variable is cross-tabulated against preference variable, the results of which are given in Table 5 below:

Table 5: CROSS TABULATION OF PREFERENCE WITH AGE GROUP

A Group / Preference	Younger Respondents	Older Respondents	Row Total
Not Interested	20 (52.63%)	13 (59%)	33
Interested	18 (47.37%)	9 (41%)	27
Column Total	38 (100%)	22 (100%)	60

The table indicates that there are 38 younger and 22 older respondents. Out of the 38 younger respondents, 47.37% are interested in the new design. However in case of olders respondents only 41% have shown an interest in favour of new design. The results indicate that the preference for the new design decreases with the age of the consumers. A plausible reason for this could be that the old customers have got used to the old design and therefore are showing resistance towards the new design.

Preference vs. Sex: In Table 1 the respondents were classified according to their sex. This variable when cross tabulated against preference variable shows the following results.

Table 6: CROSS TABULATION OF PREFERENCE WITH SEX

Preference	Not Interested	Interested	Row Total
Female	15 (55.6%)	12 (44.4%)	27 (100%)
Male	18 (54.55%)	15 (45.45%)	33 (100%)
Column Total	33	27	60

There are 27 female and 33 male respondents in the sample. Out of the 27 female respondents, 44.4% have shown an inclination towards the new design whereas 45.45% of the male respondent are in favour of new design. The difference between the proportion of male and female respondents preferring the new design does not seem to be large and therefore may be treated as insignificant.

Preference vs. Marital Status: As indicated in Table 1 the respondents were divided into two categories - Married & Single. This variable when cross-tabulated against preference variable shows the following results.

Table 7: CROSS TABULATION OF PREFERENCE WITH MARITAL STATUS

Preference / Marital Status	Not Interested	Interested	Row Total
Married	18 (40%)	12 (60%)	30 (100%)
Single	15 (50%)	15 (50%)	30 (100%)
Column Total	33	27	60

The sample comprises of equal number of married and single respondents. There are 40% of married respondents and 50% of single respondents showing interest in favour of new design. This is evident from the percentages casted row wise in the above table.

Activity 1

Suppose you have preared a cross-tabulation table between two cariables Viz family size and food expenditure How will you go about interpreting any numbers that apprar in such a table

.....

.....

.....

.....

Activity 2

A survev wss conducled in a locality to estirate the consuntion expendirure of houst holds. A sample of 150 households was taken and data on a number of variales was collected it was found that 50,30 and 20 percent of households spend less then Rs. 2000 between Rs. 2000 and 4500. and greter them Rs.4500 respeelivel on food thers per monch One of the cross-taliulaturs of the survey is shown below

	Monthly Income (Rs.)			
	< 5000	5000 - 9000	> 9000	Total
< 2000	45%	30%	25%	100%
2000 - 4500	30%	40%	30%	100%
> 4500	10%	40%	50%	100%

Have the percemages been cast in the right direetion If not why Reeast the percem ages in right direction and interprct the results

.....

.....

.....

.....

11.3 CHI-SQUARE TEST FOR ANALYSIS OF ASSOCIATION

In the previous section we discussed the association between two cross-tabulated variables by computing percentages in the direction of causal variable. However the same analysis can be carried out by using chi-square tests of independence between two variables which are categorized into two or more groups. This may be examined by testing the following hypothesis.

H_0 (Null hypothesis): The two variables are not related.

H_1 (Alternative hypothesis): The two variables are dependent.

To test the hypothesis, one applies chi-square test of independence The observed frequencies are obtained from the survey data whereas the corresponding expected frequencies are computed under the assumption that the null hypothesis is true Corresponding to a contingency table (e.g. Tables 2 to 6) the expected frequencies for any (*ij*) cell of the tables is found as

$$E_{ij} = \frac{R_i \times C_j}{G}$$

where Eij = expected frequency corresponding to the cell in ith row and jth column.
Ri = Total of observed frequencies corresponding to ith row.
Cj = Total of observed frequencies corresponding to jth column.
G = Grand total of frequencies.

The chi-square statistic is computed by using the following formula:

$$\chi^2_{(r-1)(c-1)} = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where Oij = Observed frequency of the cell in ith row and jth column.
Eij = Expected frequency of the cell in ith row and jth column.

and (r-1)(c-1) indicates the degrees of freedom where r stands for the number of rows and c for number of columns.

For any given level of significance , the computed chi-square value is compared with the tabulated chi-square value. In case computed chi-square is greater than the tabulated chi-square, we reject the null-hypothesis to conclude that the variables are dependent.

We will test the following sets of hypothesis.

1. H0. Preference for the new design is independent of education.
   H1. Preference and education are related.

2. H0. Preference for the new design is not related to the income level.
   H1. Preference for the new design is dependent on income.

3. H0. Preference for new design is independent of geographical location.
   H1. Preference for new design is related to geographical location.

4. H0. Preference for new design is not related to the age.
   H1. Preference for new design is dependent on age.

5. H0. Preference for new design is independent of sex of the respondent.
   H1. Preference for new design is dependent on sex of the respondent.

6. H0. Preference for new design is independent of marital status of the respondent.
   H1. Preference for new design is dependent on marital status of the respondent.

The tables of observed and expected frequencies corresponding to the above sets of hypothesis are given below (see Tables 8 to 13).

Table 8: OBSERVED & EXPECTED FREQUENCIES

Preference	Education		
	Under-graduation	Graduation & Above	Row Total
Not Interested	20(14.3)	13 (18.7)	33
Interested	6 (11.7)	21 (15.3)	27
Column Total	26	34	60

Note: Figures in the parenthesis ( ) represent expected frequencies.

Table 9: OBSERVED & EXPECTED FREQUENCIES			
Preference	Preference		
	Not Interested	Interested	Row Total
Poor Class	10 (6.05)	1 (4.95)	11
Middle Class	22 (22.55)	19 (18.45)	41
Upper Class	1 (4.40)	7(3.60)	8
Column Total	33	27	60

Note: Figures in the parenthesis ( ) represent expected frequencies.

Table 10: OBSERVED & EXPECTED FREQUENCIES			
Preference	Preference		
	Not Interested	Interested	Row Total
North	10 (8.25)	5 (6.75)	15
South	10 (8.8)	6 (7.2)	16
East	7 (8.25)	8 (6.75)	15
West	6 (7.70)	8 (6.30)	14
Column Total	33	27	60

Note: Figures in the parenthesis ( ) represent expected frequencies.

Table 11: OBSERVED & EXPECTED FREQUENCIES			
Preference	Age Group		
	Younger Respondents	Older Respondents	Row Total
Not Interested	20 (20.9)	13 (12.1)	33
Interested	18 (17.1)	9 (9.9)	27
Column Total	38	22	60

Note: Figures in the parenthesis ( ) represent expected frequencies.

Table 12: OBSERVED & EXPECTED FREQUENCIES			
Sex	Preference		
	Not Interested	Interested	Row Total
Female	15 (14.85)	12 (12.15)	27
Male	18 (18.15)	15 (14.85)	33
Column Total	33	27	60

Note: Figures in the parenthesis ( ) represent expected frequencies.

Table 13: OBSERVED & EXPECTED FREQUENCIES			
Marital Status	Preference		
	Not Interested	Interested	Row Total
Married	18 (16.5)	12 (13.5)	30
Single	15 (16.5)	15 (13.5)	30
Column Total	33	27	60

Note: Figures in the parenthesis ( ) represent expected frequencies.

The computed chi-square value, tabulated chi-square value, the degrees of freedom and the decision for acceptance or rejection of Hypotheses numbering 1 to 6 is summarized in the Table 14.

Table 14: COMUTED & TABULATED VALUES OF CHI-SQUARE, DEGREES OF FREEDOM & DECISION REGARDING VARIOUS HYPOTHESES				
Hypothesis Number	Chi-square computed	Degrees of freedom (r-1) (c-1)	Chi-square tabulated	Decision
1	08.910	1	3.8415	Reject H_0
2	11.600	2	5.9915	Reject H_0
3	02.440	3	7.8147	Do not reject H_0
4	00.235	1	3.8415	Do not reject H_0
5	00.006	1	3.8415	Do not reject H_0
6	00.606	1	3.8415	Do not reject H_0

By examining the results of above table we may conclude to reject Hypothesis number 1 and 2 whereas we do not have enough evidence to reject Hypithexes number 1 3, 4, 5 & 6. This shows that preference for new design is relaued to educarion and mcome level However, Preferencc for mew desogn is indepcndent of goograptucal location age sex and marital status. One precaution which may be taken computing chi swure statistic is that the observed frequency in each cell should be atlcast 5 However we have ignored this, as the objective was merely to show the compulation for chi-square test Whenrver the frequencies in any cell is less than 5, it is advised to redctine the groups in such a way that the frequency for each cell is atleast 5

Activity 3

A marketing manager was given the cross-tabulaicd data in the following table to show the assocaition between age and prefcrence for forergn brand of shals use chi-sqiarc test to examine an uppropriate hypothesis what conclusiors should be drawm

	Age		
Preference	Under 40	40 and over	Total
Interested	80	30	110
Not Interested	20	60	80
Total	100	90	190

11.4 STRENGTH OF ASSOCIATION BETWEEN TWO NOMINAL VARIABLES

Having found that there is a relationship between variables, it would be interesting to examine the strength of association between two variables There are three methods available to determine the degree of association. They are outlined as below.

1. **Contingency Coefficient:** When the null hypothesis of independence between two nominal variables is rejected, we may desire to measure the degree of assocaition between them. One of such measure called contingency coefficient (C) is given as:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

Where z2 is computed by the method discussed in previous section and n is the total size of the sample. We know that the computed value of χ^2 corresponding to hypothesis number 1 is 8.91. The corresponding contingency coefficient (C) is computed as:

$$C = \sqrt{\frac{8.91}{60+8.91}} = 0.36$$

The question which may be asked at this stage is whether a value of 0.36 for the contingency coefficient indicates a strong or weak association. It has to be compared against its limits. The lowest value of limit is zero (C = 0) when $\chi^2 = 0$ and which is possible when there is absolutely no association between variables. However for a perfectly correlated situation, this value can not be unity as is the case for linear correlation coefficient (see Unit 17 of MS-8 course). This makes the task of interpreting the value of C difficult.

However, for a contingency table, where the number of rows and columns are equal, the upper limit of C is given by $\sqrt{\frac{(r-1)}{r}}$, where r is the number of rows or columns. Since the contingency table corresponding to hypothesis number 1 has two rows and two columns, the maximum possible value of C is $\sqrt{1/2} = 0.707$.

This suggests there is a moderate association between two variables since the computed value of the contingency coefficient (0.36) is halfway between the limits of zero and the upper limit of 0.707.

As already explained, it is not possible to measure meaningfully the contingency coefficient when the number of rows and columns in the table are different. An other measure called cramer's V-statistic has been proposed under such a situaion. The expression for such a statistic is given by

$$V = \sqrt{\frac{\chi^2}{n(f-1)}}$$

where χ^2 and n are defined as before and f is the samller of the number of rows and columns in a contingency table, that is, f = min (r,c).

It may be noted that the maximum value that the computed χ^2 can take is n (f-1). Therefore V = 0, when there is lack of any association between variables and V = 1 when there is a perfect relationship. The value of V corresponding to hypothesis number 2 (where there are three rows and two columns in the contingency table) is given as

$$V = \sqrt{\frac{\chi^2}{n(f-1)}} = \sqrt{\frac{11.60}{60 \times 1}} = 0.44$$

This shows a moderate relationship between preference and the income level of the respondents.

2. **Phi Correlation Coefficient:** This is used to measure strength of association for two nominal variables in a contingency table of order 2 x 2 and is termed as phi-coefficient (ϕ). This phi-coefficient like correlation coefficient can assume any value between -1 and 1. Further, ϕ^2 (the square of phi-coeffiient) measures the proportion of one variable that is explained by the other variable.

Let us consider 2x2 cross-tabulation of Table 8 corresponding to Hypothesis number 1. We wish to examine the strength of association between preference for new design and education level. For the sake of convenience, the data is reproduced below in Table 15.

Table 15: PREFERENCE VS. EDUCATION

Preference	Education		Row Total
	Under-graduation	Graduation & Above	
Not interested	20 (a)	13 (b)	33 (a + b)
Interested	6 (c)	21 (d)	27 (c + d)
Column total	26 (a + c)	34 (b+d)	60 (a+b+c+d)

Phi-coefficient (ϕ) may be computed by using the following formula:

$$= \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$= \frac{(20)(21) - (13)(6)}{\sqrt{(33)(27)(26)(34)}}$$

$$= \frac{342}{\sqrt{787644}}$$

$$= \frac{342}{887.49} = 0.385$$

The following description of the strength of relationship for a given particular phi value may be used.

Value of $\pm \phi$	Strength of Relationship
Greater than 0.80	Strong
0.40 to 0.80	Moderate
0.20 to 0.40	Weak
0.00 to 0.20	Negligible

Source: Marketing Research by Luck & Rubin p-503.

According to the above description, there is a weak association between preference and education. The value of $\phi^2 = 0.15$ indicates that 15% of the variations in preference is explained by education level.

The phi-coefficient may assume positive or negative value. However, sign of ϕ does not have any particular meaning. If the responses were concentrated in cells c & b instead of a and d, the sign of phi would have been negative.

It may be noted that when total sample size increases, the computing formula for phi may become tedious. In such a situation, the phi-coefficient may be computed as

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

3. **Goodman and Kruskal's Lambda:** The two measures namely contingency coefficient and phi are the symmetrical measures of association in the sense that they do not predict which variable predicts the other. However, there can be instances where a marketing researcher may desire to determine the direction of prediction. To meet such a requirement Goodman and Kruskal's Lambda - asymmetric coefficient may be used. This method measures the extent to which we may reduce the error in predicting categories of one variable from the knowledge of the categories of other variable.

Let us use the cross-tabulation given in Table 2. Suppose we wish to predict preference from education level We would need to rearrange Table 2 which may appear as:

Table 16: CROSS TABULATION OF PREFERENCE WITH EDUCATION

Preference / Education	Not Interested	Interested	Row Total
Undergraduation	20	6	26
Graduation	13	21	34
Column Total	33	27	60

$$\lambda asym = \frac{\sum_{i=1}^{R} f_{ir} - F_c}{n - F_c}$$

where fir = the maximum frequency found within each category of the row variable.

Fc = the maximum frequency among the marginal totals of the column variable

n = sample size

R = number of categories in row variable

$\sum_{i=1}^{R} f_{ir} = 20 + 21 = 41$

Fc = 33

n = 60

$$\lambda asym = \frac{41-33}{60-33} = 0.259$$

If we want to compute λasym for the case when we want to predict education from preference, we would need to use Table 2 as our data input. We could compute

$\sum$fir = 20 + 21 = 41

Fc = 34

n = 60

Therefore, λ asym for predicting education from the preference is

$$\lambda asym = \frac{41-34}{60-34} = 0.269$$

The result indicates that depending on which variable predicts another variable, we will have different values of lamda.

If we are only interested in measuring the association (and not direction) between two variables, we may compute λ symmetrical as given by

$$\lambda sym = \frac{\sum_{i=1}^{R} f_{ir} + \sum_{i=1}^{C} f_{ic} - (F_r + F_c)}{2n - (F_r + F_c)}$$

Where fic = the maximum frequency found within each category of the column.

Fr = the maximum frequency among the marginal totals of the row variable.

C = number of categories in a column.

Applying the formula to the data in Table 2, we get

$\sum$fir = 20 + 21 = 41

$\sum$fic = 20 + 21 = 41

Fr = 33

Fc = 34

$$\lambda sym = \frac{41+41-(33+34)}{120-(33+34)} = \frac{15}{53} = 0.28$$

The term λ sym is used to represent an average predictability between the two variables of interest. Both the forms of lambda can be tested for statistical significance. However their discussion is beyond the scope of this unit.

Activity 4

State whether the following staements are true (T) or False (F).

S.No.	Statement	True/False
1	The contingenov coeltocoenl may be usrd to dekerminc the direction of predociton	
2	The upper lumit of contngenev coefficint is unuty	
3	Phi-coeflicient is computed for contngencv tables of order 2x2	
4	Phi-coefficient is a symmetrical measure of assocration between two vriables	
5	Goosman ans Kruskal s lambda mav also be used for measuring degree of assoctafion between variables	

Activity 5

List out the limitations of contingencv coefficient in interpreitng the results of assiciation between nominal variables

.....

.....

.....

.....

11.5 CORRELATION COEFFICIENT

The degree of association between two variables is also computed using correlation coefficient. A detailed treatment of correlation coefficient was taken up in Unit no. 18 of course (MS-8) Quantitative Analysis for Managerial Application. We know that the correlation coefficient may be positive, negative or zero. A measure of linear correlation coefficient between two variables X & Y is measured by correlation coefficient, the formula of which is given below:

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2}\sqrt{\Sigma(Y-\bar{Y})^2}}$$

Where $\bar{X}$ = Sample mean for the variable X

$\bar{Y}$ = Sample mean for the variable Y

n = Number of observations in the sample

This measures can assume any value between -1 and +1 and is independent of units of measurements. The correlation coefficient can be computed for interval and ratio scale data. If we examine the data in Table 1, we find that preference data indicated in column 2, age data in column 6 and acutal income data in column 8 can be utilised to compute correlation coefficient. For the sake of illustration let us compute the correlation coefficient between preference and actual income using the above mentioned formula. It is found that the correlation coefficient between preference and income is equal to .4597 indicating a positive relationship between these two variables. Further the magnitude of correlation coefficient suggests that there is a moderate relationship between preference and income

To test the significance of population correlation coefficient (ρ) we may use t test described below

H_0 (Null Hypothesis) — There is no correlation between preference and income ($\rho = 0$)

H_1(Alternative Hypothesis) — There is correlation between preference and income ($\rho \neq 0$).

The test statistic used to test the hypothesis is given by:

$$t_{n-2} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Where r = Sample correlation coefficient

n = Sample size

The significance of correlation coefficient was tested by using above formula and the computations are as given below.

$$t = \frac{0.46\sqrt{58}}{\sqrt{1-.2116}}$$

$$= \frac{0.46 \times 7.616}{0.8879}$$

$$= \frac{3.50336}{0.8879} = 3.9459$$

For a given level or signiticance=0.5 (say), the table value of t is obtained as 1.96. this calue being less than the compured value of t (3.9459), we reject the hypothesis of no correlation between the preference and actual income

Activity 6

What the differences in two measures viz phi-coefficient and coffelation coefficient in measuring the strength of association between two variables

.....

.....

.....

.....

Actovoty 7

consider the data given in Table 1. Use the preterence data in column 2 and age data in column 6 of the tabl to find if there is any correlation coefficient between these vari-ables also tesr for its significance

.....

.....

.....

.....

11.6 SIMPLE LINEAR REGRESSION

One of the lurulations of correlation analysis is that it only measures the degree of association between two varavles and does not deal with cause and effect relationship between variables regression amalysis over comes this problem by assuming a causal relationsship betweer ourables regrcssion analysis ɔnly assumes cause and effect relationship and does not esrablish ot) Assuming Y is a function of X, we are treating X as a causal variatce ane y as a detendent cariavle This means it is only the variable X

which influences the variable Y and not the other way round. We have taken a very simplistic case of a situation where the dependent variable Y is explained only by one independent (causal variable) X. In real life situation there may be a host of variables influencing the dependent variable Y. The discussion of such a case comes under multiple regression model and will be taken up in Unit no. 12 of this course. Here our concern is with simple linear regression model where there is only one dependent and one independent variable. Assuming a linear relationship between the dependent and the independent variable, the relationship between Y and X may be written as

$$Y = b_0 + b_1 X + u$$

where b_0 = intercept term

b_1 = slope term

u = random error

The estimation of b_0 and b_1 is carried out by using ordinary least squares method of estimation (for details on this method, please consult Unit no. 19 of MS - 8 Course.) The ordinary least squares method (OLS) aims at minimizing the error sum of squares to obtain the estimated value of b_0 and b_1. The estimated values are obtained using the following formulae:

$$\hat{b}_1 = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sum (X - \overline{X})^2}$$

$$\hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{X}$$

where $\overline{X}$ = Sample mean of the variable X

$\overline{Y}$ = Sample mean of the variable Y

$\hat{b}_0, \hat{b}_1$ = Estimates of population parameters b_0 and b_1

We have taken preference as the dependent variable and income as an independent variable and have run a simple regression of preference on income. The results are given below:

Y	=	1.79 + 0.00027X	
t		(3.94)	(4.68)

The above estimated regression equation indicates that as the income increases by one rupee the preference reading goes up by .00027 points. Further the positive sign of the slope coefficient indicates that the relationship is positive. The significance of the slope cofficient is tested with the help of t statistic. Here the computed value of t is 4.68 and if we assume 5% level of significance and two tailed alternative, the tabulated value of t would be 1.96. It is seen that the computed t is greater than the tabulated t indicating that the cofficient corresponding to income variable is significant at 5% level.

The goodness of fit of the regression equation is estimated by r^2, the cofficient of determination. The value of r^2 is always between 0 and 1 and higher the value of r^2 better is the goodness of fit. The value of r^2, also represents the explanatory power of the regression model. If the value of r^2 equals .84 (say), it means that 84% of the variations in the dependent variable are explained by the independent variable. In the case of our example the value of r^2 equals .21 which means 21% of the variation in the preference variable are explained by income variable. (We have very briefly touched upon this section. For details on regression analysis please refer to Unit no.19 of MS-8 Course.)

Activity 8

Consider the data on preference(column2) and agc (colmn6) variables as given in Tuble l Comute the simple regression of preference for new desogn on age and carry out appropriate statical test of sognificamce Also interpet your results

.....

.....

.....

.....

11.7 ANALYSIS OF DIFFERENCES

In this sectirn we would like to answer some of the questions such as

1 Is there any difference betwen the monthly inconrne of respondents who are interested in the new desogn form those who are not interested in the new design

2 Is there any defferance in the proportion of female and male respondents preferring the new desogn

The above question may be answered by testing the following hvpotheses

1. H_0 (Nuil Hypothesis)	There is no difference befween the monthly income of those who are interested in the new design and those not interesed in the new design ($\mu_1=\mu_2$)
H_1 (Alternative Hypothesis)	The monthly income of the respondents preferring the new design is not eual to those not preferring the desogn ($\mu_1=\mu_2$)
2. H_0 (Null Hypothesis)	The proportion of females preferring the new design is equal to the proportion of males preferring the new design ($\pi_1=\pi_2$)
H_1 (Alternative Hypothesis)	the proportion of females preferring the new design is not equal to the proportion of males preferring the new desin ($\pi_1=\pi_2$)

Where

μ_1 = Population mean income of those preferring the new design

μ_2 = Population mean income of those not preferring hte new sesogn

π_1 = proportion of females in the [opulation preferring the new sesign

π_2 = proportion of males in the population preferring the new design

Tests of Difference between Means

To test the difference between two population means one needs to apply either tor z test . Appendix 1 of this unit discusses the situation where each of these of these tests are applicable In our present case are 27 respondents preferring the new design (n_1=27)and 33 respondents(n_2=33) not preferring the new design the vaeiable of interest is income For testing the Hypothesis number 1 we have computed the following

$\bar{X}_1$	=	6,218.6665
$\bar{X}_2$	=	4,277.4243
s_1	=	1,885.9080
s_2	=	1,830.6273
Standard Error of difference between Means $\left(\hat{\sigma}_{\bar{X}_1 - \bar{X}_2}\right)$	=	481.5306
Computed t	=	4.03
Critical t with 58 d.f. (assuming 1% level of significance)	=	2.576

Since computed t is greater than tabulated t, we reject the null hypothesis and conclude that the monthly income of respondents preferring the new design is different from those not preferring the new design.

Tests of Difference Between Proportion

The difference between two population proportions may be tested using z test, assuming normal approximation to binomial distribution. Appendix 1 of this unit discusses the test briefly. For testing Hypothesis 2, we have 27 females (n_1 = 27) and 33 male respondents (n_2 = 33). The variable of interest is the proportion of respondents preferring the new design. In our case we have the following computed information.

Proportion of females preferring the new design (p_1)	=	0.4444
Proportion of males preferring the new desing (p_2)	=	0.4545
Estimated standard error of difference between proportion ($\hat{\sigma}_{P_1 - P_2}$)	=	0.1314
Computed z value	=	0.077
Critical z value at 5% level of significance	=	1.96

Since computed z value is less than the critical z value we accept the null hypothesis of no difference between two population proportion. Therefore the proportion of females preferring the new designs is not significantly different from the proportion of males preferring the new design.

Activity 9

Consider the data on televant variables to test the hypothesis that there is on difference in the aveerage age of those who prefer the new dcsign and the one who do not . Yiu may use 5%level of sugbufucance

.....

.....

.....

.....

11.8 SUMMARY

In this unit, our concern was with bivariate analysis of data. We examined the simultaneous relationship between two variables at a time. We discussed how to interpret a cross-tabulation table by computing percentages. The rule devised was to cast the

percentages in the direction of causal variable while interpreting cross tabulation tables. We also discussed chi-square test for examining the independence between two nominally scaled variables. The strength of association between two variables was studied using three methods namely Contingency Coefficient, Phi Correlation Coefficient and Goodman & Kruskal's Lamda. The limitation of contingency coefficient is that it is not possible to measure meaningfully the contingency coefficient when the number of rows and columns in the contingency table are not equal. Another statistic called Cramer's V-Statistic is used in such a situation. The phi-coefficient is used for 2x2 table and can take values between -1 & +1. Goodman & Kruskal's Lamda are of two types viz. λasym & λsym. The first one is used to determine the direction of prediction whereas the second may be used for measuring association between two nominal variables.

The degree of two association between two variables is also computed with the help of simple correlation coefficient which may take any value between -1 & +1. However, the correlation coefficient does not deal with cause and effect relationship and to overcome this regression analysis is used. We have discussed the estimation and interpretation of regression results.

The analysis of difference between two **population means can be fcsted with the help of** t or z statistic. We have discussed the situations under **which these tests are applicable** The difference between two proportions **mau be tested with the help of z statistic**

All the above methods are **illustrated with the help of survey data exhibited in table l of** this unit.

11.9 SELF-ASSESSMENT QUESTIONS

1. What are the ways in which **percentages may be computed in a cross-tabulation** table? which way is the best?
2. In a contingency table, what **type of hypothesis is testcd using chi-square test** What precautions **may be taken while applying the test**
3. What are the **various measures usrd to measure the strength of assoeiation** between two **nominal variables ? Describe them and clearly mention their lumitations** if any.
4. Discuss the problem of using the **contimgency coeffocient in inteorering the results** of correlation analysis?
5. Explain **difference between corretation & regression**
6. A marketing **rescatcher intested in the business publication reading habits of** purchasing agents has **asscmbled the following data in the form of cross-tabulation** alongwith the **information whether each agent holds a degree or not**

	First choice mentions		
Business Publication	**Degree**	**No degree**	**Total**
W	20	15	35
X	15	15	30
Y	25	20	45
Z	30	25	55
Total	90	75	105

a) Is there an **assoviation betwcem business-publication choice and type of** college degree?

b) What is the appropriate **null hypothesis for this illusuation**

c) Use an appropriate **method to mcasuire the stregth of association between** two variables.

7. A marketing research anlyst has one nominally scaled variable and one interval scaled variable. The analyst wishes to use a chi-square test to examine the association between two variables. Give your suggestion as to what can be done.

8. What type of scalar data (Nominal, Ordinal, Interval or Ratio) is typically utilized in cross tabulation analysis?

9. A new breakfast cereal is being test marketed in selected cities on the east and west coasts. Consumer panels are being used for the evaluation in each of the selected cities, and after four weeks of product use, the consumer reactions have been obtained as follows:

	East coast	West coast
Total responses	632	428
Preferred cereal or considered cereal equivalent to others	468	327

We are interested in finding out whether or not there are regional differences in consumer acceptance of the product. Conduct an appropriate test on the results of two samples at 1% level of significance.

10. Two batches of the same product are tested for their mean life. Assuming that the lives of the product follow a normal distribution with an unknown variance, test the hypothesis that the mean life is the same for both the batches, given the following information:

Batch	Sample Size ($\bar{X}$)	Mean life in hrs. (s)	Standard deviation
I	10	750	12
II	8	820	14

Choose a significance level of 5%.

11. An automobile manufacturing firm is bringing out a new model. In order to map out its advertising campaign, it wants to determine whether the model will appeal most to a particular age-group or equally to all age-groups.

The firm takes a random sample from persons attending a preview of the new model and obtained the results summarised below:

	Age Groups			
Persons	**Under 20**	**20 - 39**	**40 - 59**	**60 and over**
Liking the car	146	78	48	28
Disliking the car	54	32	32	62

What conclusion would you draw from the above data? You may use 1% level of significance.

12. A study was conducted to know whether there is any difference between the monthly income of the probationary officers in public sector banks to that of private sector banks. Some sample results are given below:

Sample Statistic	Public Sector Banks	Private Sector Banks
Sample size	45	60
Sample mean	Rs.9,500	Rs.10,800
Sample variance	Rs.438	Rs.725

Test the hypothesis that the income of probationary officers in private sector banks is higher than that of public sector banks using 5% level of significance.

11.10 FURTHER READINGS

Beri, G. C. "Marketing Research - Text and Cases" Tata McGraw-Hill Publishing Co. Ltd. (1st Edition).

Green, Paul E. and Donald S. Tull "Research for Marketing Decisions" Prentice-Hall of India Pvt. Ltd. (4th Edition).

Kinnear, Thomas C. and James R. Taylog, "Marketing Research - An Applied Approach" McGraw-Hill International Editions (3rd Edition).

Luck, David J. and Ronald S. Rubin, "Marketing Research" Prentice-Hall of India Pvt. Ltd. (7th Edition).

Majumdar, Ramanuj "Marketing Research - Text, Applications and Case Studies" Wiley Eastern Ltd. (1st Edition).

Westfall, Boyd and Stasch. "Marketing Research - Text and Cases" Richard D. Irwin. Inc. (6th Edition).

APPENDIX - 1

We would briefly review the statistical tests of differences between two means & proportions. It would be of help if you go through Unit no. 15 of the course on Quantitative Analysis for Managerial Applications (MS-8).

Test of two means: We want to test whether or not the observed differences between two sample means is significant. In other words - are population means really different?. To examine this hypothesis we may have to make a choice between Z and t test as was done in the case of univariate analysis. We choose Z test when standard deviations are known for both population or when sample sizes are sufficiently large (i.e., $n_1 > 30$ and $n_2 > 30$). The use of t test is made when population standard deviation for either population is unknown or when one or both samples are small (i.e. $n_1 < 30$ or $n_2 < 30$).

Z test for difference between two means

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$ (Assuming a two-tailed alternative)

Test statistic

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

Where $\bar{X}1$ = sample mean for the first variable

$\bar{X}2$ = sample mean for the second variable

μ_1 & μ_2 = hypothesized population means for two variables

$\sigma_{\bar{X}_1 - \bar{X}_2}$ = standard error of the difference between means

The standard error of difference between means may be computed as

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Where σ_1^2 *and* σ_2^2 are the population variances for the two variables and n_1 and n_2 are the respective sample sizes.

For a given level of significance α, the corresponding Z value is obtained from standard

normal tables and is compared with computed Z value. If computed Z value (in absolute term) exceeds corresponding tabulated Z value, we reject the null hypothesis (H0) of no difference between two population means.

In case σ_1 and σ_2 are unknown but both n_1 and n_2 are greater than 30, we may still use Z test. However s_1 and s_2 (sample standard deviaiton) are used to estimate σ_1 and σ_2 respectively and the estimate of standard error of difference between means is obtained as

$$\hat{\sigma}_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

where $$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2$$

and $$S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2$$

.the t test on difference between two means when one or both sample sozes are small(i.e n_1 <30 or n<30)a t- test would be approproate for usong t-test following assumptions are required

i) Independent samples are drawn from two normal populations.

ii) Population variances are unknown but equal.

The hypothesis may be stated as:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$ (Assuming a two-tailed alternative)

The test statistic is

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

with $n_1 + n_2 - 2$ degrees of freedom, where sp is an estimate of pooled standard deviation, given by

$$sp = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

For a given level of significance , the corresponding t value with n_1+n_2-2 degrees of freedom is obtained from t-distribution table. If computed t (absolute value) exceeds tabulated t, we reject the null hypothesis.

Test of two proportions: Suppose we are interested in testing the hypothesis involving the equality of two population proportions and the two sample sizes n_1 and n_2 from the populations are large enough so as to have each of the quantities n_1p_1, $n_1(1-p_1)$, n_2p_2, and $n_2(1-p_2)$ atleast 5, we may use normal approximation to binomial distribution to apply Z test. The hypothesis to be tested is:

$H_0 : \pi_1 = \pi_2$

$H_1 : \pi_1 \neq \pi_2$ (Assuming a two-tailed test)

The appropriate test statistic is

$$Z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sigma_{p_1 - p_2}}$$

Where p_1, p_2 = sample proportions

π_1, π_2 = hypothesized population proportion

$\sigma_{p_1 - p_2}$ = population standard error for the difference between proportions

An estimate of the population standard error for the difference between proportions is given by

$$\hat{\sigma}_{P_1 - P_2} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where $\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

$$\hat{q} = 1 - p$$

For a given level of significance α, the value of Z is obtained from standard normal table. If computed value of Z (absolute value) is greater than the corresponding tabulated value, we reject H_0.

Notes

Notes:

Uttar Pradesh
Rajarshi Tandon Oprn University

Block

4

MULTIVARIATE ANALYSIS

UNIT 12
Regression Analysis, Discriminant Analysis and Factor Analysis 5

UNIT 13
Conjoint Analysis 27

UNIT 14
Cluster Analysis and Multi-Dimensional Scaling 36

UNIT 15
Application of Marketing Research in India — Some Case Studies 47

Expert-Cum-Course Writers Committee

Prof. Rakesh Khurana, *Chairman*
School of Management Studies
IGNOU, New Delhi

Prof. Abraham Koshy
Indian Institute of Management
Ahmedabad

Prof. Y.L. Grover
Management Consultant
New Delhi

Prof. Deepak Chawla
International Management Institute
New Delhi

Prof. Mukul Gupta
International Research Bureau
New Delhi

Mrs. Neerja Wable
Indian Market Research Bureau
New Delhi

Shri Saurabh Khoshla
MD, Tulika Advertising Agency
New Delhi

Prof. Sudas Roy
Indian Institute of Management
Calcutta

Dr. N Kinra
Indian Institute of Management
Lucknow

Prof. A.P. Arora
Management Development Institute
Gurgaon

Dr. Sanjay K. Jain
Delhi School of Economics
University of Delhi, Delhi

Mrs. Monica Khanna,
New Delhi

Dr. P.S. Vishwanathan
Hindustan Computer Ltd.
New Delhi

Prof. Sharad Sarin
XLRI
Jamshedpur

Prof. Mohan Aggarwal
XLRI
Jamshedpur

Mr. D.K. Bose
Media Director
R.K. Swamy, BBDO, New Delhi

Dr.(Mrs.) Madhulika Kaushik
Director
School of Management Studies
IGNOU, New Delhi

Prof. R.K. Grover
School of Management Studies
IGNOU, New Delhi

Mr. Qasemuddin Haider
Asst. Regional Director
IGNOU, Patna

Mr. Ravi Shanker
School of Management Studies
IGNOU, New Delhi

Dr. Kamal Yadava
Course Coordinator
School of Management Studies
IGNOU, New Delhi

Print Production: B Natarajan, Copy Editor, SOMS, IGNOU

July, 1998 (Reprint)
Indira Gandhi National Open University, 1998
© ISBN-81-7605-367-8
All rights reserved. No part of this work may reproduced in any form, by mimeograph or any other means, without permission in writing from the Indira Gandhi National Open University. Further information of the Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi - 110 068.

Reproduced and reprinted with the permission of Indira Gandhi National Open University by G. S. Shukla, Registrar, U.P.R.T.Open University, Allahabad (December 2011)
Printed by : Chandrakala univrsal put. Ltd. 42/7 Jawahar Lal Nehru Road Allahabad 211002

UGSEC-02/206

BLOCK 4 MULTIVARIATE ANALYSIS

This block gives you the details of various multivariate analysis techniques which can be used for marketing research. The block has four units. Unit 12 explains the techniques of regression analysis, discriminant analysis and factor analysis. Unit 13 gives the details of conjoint analysis - a multivariate technique which was introduced to marketers in 1970's and which has gained tremendous popularity since then due to its great managerial relevance and simplicity.

Unit 14 covers two techniques viz. cluster analysis and multi dimensional scaling. The last unit of the block gives some case studies on applications of marketing research in India.

UNIT 12 REGRESSION ANALYSIS, DISCRIMINANT ANALYSIS AND FACTOR ANALYSIS

Objectives

After studying this unit, you should be able to .

- explain the concept of Association that takes place between a dependent variable and a set of independent variables
- describe the various multivariate procedures available to analyse associative data in the context of marketing
- interpret the findings of multivariate analysis in any market research study
- use a particular technique of multivariate analysis suitable for a particular marketing problem.

Structure

12.1 Introduction

12.2 Analysis of Variance

12.3 Regression Analysis

12.4 Discriminant Analysis

12.5 Factor Analysis

12.6 Summary

12.7 Self-Assessment Exercises

12.8 Further Readings

12.1 INTRODUCTION

In the previous block we covered the fundamentals of statistical inference with special emphasis on hypothesis testing as an effective tool for marketing decisions. Univariate analysis forms the foundation for the development of multivariate analysis, which is the topic of discussion in this unit. While the concept of the univariate analysis will continue to draw our attention time and again, our focus in this unit will be on procedures of multivariate analysis which has emerged as the most striking trend in marketing research methodology.

Description and analysis of associative data involves studying the relationship and degree of association among several variables and therefore multivariate procedures become imperative. We shall attempt to highlight the procedures with a marketing orientation. It is important to realise that amongst the many techniques available, some are more intensely used compared to others.

Analysis of variance is suitable for analysing experimental data based on field experiments which is gaining increased attention in marketing research. It is a useful analytical tool for marketing researchers. Whenever we are interested in comparison of means of any number of groups, analysis of variance is an appropriate technique to use. This technique could be applied in new product evaluation, selection of copy theme, effectiveness of an advertising/sales promotional campaigns and the like.

Regression analysis finds out the degree or relationship between a dependent variable and a set of independent variables by fitting a statistical equation through the method of least square. Whenever we are interested in the combined influence of several independent variables upon a dependent variable our study is that of multiple regression. For example demand may be influenced not only by price but also by growth in industrial production, extent of import prices of other goods, consumer's income, taste and preferences etc. Market researchers could use regression for

explaining per cent variation in dependent variable caused by a number of independent variables and also problems involving prediction or forecasting.

Discriminant analysis is useful in situations where a total sample could be classified into mutually exclusive and exhaustive groups on the basis of a set of predictor variables. Unlike the regression analysis, these predictor variables need not be independent. For example one may wish to predict whether sales potential in a particular marketing territory will be 'good' or 'bad' based on the territory's personal disposal income, population density and number of retail outlets. You may like to classify a consumer as a user or non-user of one of the five brands of a product based on his age, income and length of time spent in his present job. Here the interest is what variables discriminate well between groups.

Factor analysis provides an approach that reduces a set of variables into one or more underlying variables. The technique groups together those variables that seem to belong together and simultaneously supplies the weighing scheme. For example one may be interested in the identification of factors that determine the company's image. When the decision maker is overwhelmed by the factor analysis comes to his help in compressing there many variables into a few meaningful dimensions, like service orientation, quality level and width of assortment in a research project involving 20 retail chains on 35 factors or variables.

12.2 ANALYSIS OF VARIANCE

Analysis of variance abbreviated as ANOVA is the hallmark of statistical methodology pioneered by R.A. Fisher, while analysing design of experiments. This technique is applicable for analysing experimental data. Variation is inherent in nature. The analysis of variance technique attempts to find out how much variation is caused by random fluctuations called errors and how much by assignable causes. The technique breaks down the total variation into meaningful components variation — that produced by treatments and that produced by random errors. Here the word 'treatment' implies exposing one set of experimental data to a particular device used in an experiment. The group or set receiving this treatment is called a control group. For example target consumers could be put into two groups, one group is asked to taste the new brand of tea while the other group is asked to taste the existing brand; or each group is asked to taste first the existing brand in the morning and the new brand in the evening. The objective could be to evaluate whether there is any significant difference in the perception by the two groups.

Whenever we are interested in the comparison of two or more treatment means, the most appropriate tool is the analysis of variance. For example, a test to measure the sales appeal effectiveness of three different package designs or a test to measure the elasticity of four different prices, could be carried out through analysis of variance method. If only one test variable is used then it is called one way analysis of variance. If two test variables are used then it is called two way analysis of variance. The discussion will begin with one way analysis of variance which lays the foundation for two way analysis of variance. We will avoid complicated mathematical derivations and focus only on procedure and hypothesis testing relevant to marketing research. However for the assumption underlying the model and computational aspect, certain mathematical rotations and algebra become unavoidable. One thing is sure that you will be very clear when the illustrations are discussed.

One way classification

ANOVA Model $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$

$i = 1, 2, \ldots\ldots, a$

$j = 1, 2, \ldots\ldots, n_i$

where Y_{ij} is jth observation corresponding to ith treatment drawn independently.

a is the number of treatments

n_i is the number of observations on each treatment

$\sum_{i=1}^{a} \alpha_i = 0 \quad \epsilon_{ij} \cap IND\ (0, \sigma^2)$

μ is the grand mean effect

α_i is the ith treatment effect

IND means independently normally distributed. It can be shown by minimising the sum of squares of errors with respect to μ, α_i, and then squaring

Total sum of squares
= Treatment sum of squares + Error sum of squares

Computational aspect :

Correction factor $C = \frac{G^2}{n}$ where G is the grand total of all observations, n is the total number of observations $= \sum_{i=1}^{a} n_i$

Total Sum of Squares (TSS) = TSS = $\sum_{i,j} y_{ij}^2 - C$

Treatment Sum of Squares (TRSS) = TRSS = $\frac{\Sigma Y_i^2}{ni} - C$

Error Sum of Squares (ESS) = TSS – TRSS

Alternative ESS can also be computed directly as

$$\sum_{ij} Y_{ij}^2 - \sum_{i} \frac{Y_i^2}{r_i}$$

This would help cross checking the result obtained by substraction as before (TSS – TRSS)

Y_i = Marginal total of ith treatment
i = 1, 2, ..., a

Form the ANOVA Table as follows :

Source of variation	D.F. (Degrees of Freedom)	Sum of Squares	Mean Square	F (ratio)
Due to Treatment	a–1	TRSS	TRSS/a – 1	(TRSS/a – 1). (ESS/n – a)
Due to Error	n – a	ESS	ESS/n – a	
Total	n – 1	TSS		

$$\frac{TRSS/(a-1)}{ESS/(n-a)} = \frac{\text{Treatment Mean Square}}{\text{Error Mean Square}}$$

Follows Sredecor's F distribution with (a – 1), (n – a) d.f.

In symbol $\frac{TRSS/(a-1)}{ESS/(n-a)} \sim F(a-1, n-a)$

Set up H_0 : Treatment Means are equal
H_1 : There is difference amongst treatment means.

If the calculated F in the ANOVA table exceeds the Table F (a–1, n–a) at 5% level reject H_0 and accept H_1.

Easy way to remember computational aspect of ANOVA.

1) If there are a treatments (generally columns) then degrees of freedom for treatment = a–1
2) If the total number of observation is n degrees of freedom associated with Total = n–1
3) Degrees of freedom associated with error = (2) – (1)
   = (n–1) – (a–1)
   = n–a
4) Sum all observations, square it and divide by n, you get correction factors C.
5) Square each observation; sum them and subtract C, you get total sum of square = TSS

6) Obtain the marginal total for each treatment i (column) square it and divide by n_i the number of observations under column i. Sum for all i and subtract c, you get the Treatment Sum of Squares = TRSS

7): (5) – (6) gives error sum of squares = ESS

Now you should be in a position to form the ANOVA Table.

To make matters clear let us look at an example from marketing.

A consumer product company was interested in evaluating the sales effect of two different colours for the package of one of its products. The firm selected 10 stores with similar monthly sales pattern, and randomly split them into two groups of 5 stores each. One group of stores was stocked with red colour packages while the other group of stores was stocked only with blue packages. All stores were monitored for two weeks to make certain that the packages were properly displayed and that no stockout occurred. The following data shows the number of test packages which were sold in each store for the two-week period. The company wants to know whether there is significant difference in sales effectiveness of the two packages.

Sales by stores

Red package	Blue package
6	16
8	18
10	20
12	22
14	24

Computation :

There are 2 treatments here — Red package and Blue package. $n_1 = 5$ (number of observations under treatment 1) and $n_2 = 5$ (number of observations under treatment 2 $n = n_1 + n_2 = 5 + 5 = 10$

	Treatment 1	Treatment 2	Total
	6	16	22
	8	18	26
	10	20	30
	12	22	34
	14	24	38
Total	50	100	150
	G = 150.		

$$\text{Correction factor} = \frac{G^2}{n} = \frac{150^2}{10} = \frac{22500}{10} = 2250$$

$$TSS = 6^2 + 8^2 + 10^2 + 12^2 + 14^2 + 16^2 + 18^2 + 20^2 + 22^2 + 24^2 - 2250 = 330$$

$$TRSS = \frac{50^2}{5} + \frac{100^2}{5} - 2250 = 250$$

$$ESS = TSS - TRSS = 330 - 250 = 80$$

Anova Table

Source of variation	D.F.	Sum of square	Mean squares	F Ratio
Due to treatment	1	250	$\frac{250}{1} = 250$	$\frac{250}{10} = 25$
Due to error	8	80	$\frac{80}{8} = 10$	
Total	9	330		

Null Hypothesis H_0 : There is no difference between red package and blue package in sales effectiveness.

H_1 : There is difference between red and blue package in regard to sales effectiveness.

Table F (1, 8) at 5% level = 5.32

Interpretation of results

Since the calculated F = (25) exceeds table F (1, 8) = 5.32, reject the null hypothesis that the treatment means are equal and accept H_1. This means at 5% level or with a confidence level of 95%, we can state that there is difference in sales effectiveness between red package and blue package. The analysis of variance does not tell the researchers which of the two package is more effective. By observation, we can conclude that the blue package was more effective than the red package.

Two-way analysis of variance

The basic procedures underlying one way analysis of variance apply also to two way analysis of variance, except for the fact that in two-way analysis of variance, there are two variables being tested rather than only one.

The model :

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

The assumptions and structure are as before except that we have added one more factor β_j for which $\Sigma \beta_i = 0$. α_i denotes treatment or column effect β_j denotes block or row effect. The computational aspects are same as before except that we have to calculate row sum of squares (or block sum of squares). To recapitulate let us write down the expression.

$$\text{Correction factor } C = \frac{G^2}{n}$$

$$\text{Total sum of Squares (TSS)} = \sum_{i,j} Y_{ij}^2 - C$$

$$\text{Treatment Sum of Squares (TRSS)} = \sum_{i} \frac{Y_i^2}{n_i} - C$$

$$\text{Block Sum of Squares (BSS)} = \sum_{j} \frac{Y_j^2}{a} - C$$

$$\text{ERSS} = \text{TSS} - \text{TRSS} - \text{BSS}$$

BSS involves marginal total for jth block or row, each containing number of observations = a being the number of treatments in each block.

ANOVA Table for two-way classifications

Source	D.F.	Sum of squares	Mean square	F Ratio
Due to Treatment	a–1	TRSS	$\frac{TRSS}{a-1}$	$\frac{(TRSS)}{(a-1)} \div \frac{(ESS)}{(n-a-b+1)}$
Block	b–1	BSS	$\frac{BSS}{b-1}$	$\left(\frac{BSS}{b-1}\right) \div \left(\frac{ESS}{n-a-b+1}\right)$
Error	n–a–b+1	ESS	$\frac{ESS}{n-a-b+1}$	
Total	n–1	TSS		

obtain by subtraction from total degrees of freedom, degree of freedom for treatment and block you please note that there are two F Ratios in this ANOVA, one to treatment effect and one to block effect.

H_0 : Treatment Means are equal
H_1 : Treatment Means are not equal } For treatment

H_0 : Block Means are equal
H_1 : Block Means are not equal } For Block

If the calculated F exceeds Table F at 5%, reject H_0 and accept H_1.

Example

The following table gives the quality ratings of 10 service stations by five marketing research professionals. What are your conclusions on service effectiveness difference and difference in professional.

	Service Stations										
Raters	1	2	3	4	5	6	7	8	9	10	Total
A	99	70	90	99	65	85	75	70	85	92	830
B	96	65	80	95	70	88	70	51	84	91	790
C	95	60	48	87	48	75	71	93	80	93	750
D	98	65	70	95	67	82	73	94	86	90	820
E	97	65	62	99	60	80	76	92	90	89	810
Total	485	325	350	475	310	410	365	400	425	455	4000

Computations

$$\text{Correction Factor } C = \frac{G^2}{n} = \frac{4000^2}{50} = 320000$$

$$TSS = 99^2 + 96^2 + 95^2 + 98^2 + 97^2 + \ldots\ldots\ldots + 92^2 + 91^2 + 93^2 + 90^2 + 89^2 - 320000 = 9948$$

$$TRSS = \frac{485^2}{5} + \frac{325^2}{5} + \ldots\ldots + \frac{455^2}{5} - 320000 = 6810$$

$$BSS = \frac{830^2}{10} + \frac{790^2}{10} + \frac{750^2}{10} + \frac{820^2}{10} + \frac{810^2}{10} - 320000 = 400$$

$$ESS = 9948 - 6810 - 400 = 2738$$

ANOVA Table

Source	D.F.	Sum of Squares	Mean Square	F Ratio
Treatment (Column)	9	6810	$\frac{6810}{9} = 756.67$	$\frac{756.67}{76.06} = 9.95$
Block (Row)	4	400	$\frac{400}{4} = 100.00$	$\frac{100.00}{76.06} = 1.31$
Error	36	2738	$\frac{2738}{36} = 76.06$	
Total	49	9948		

Treatment (column) variations :

Null hypothesis H_0 : There is no difference between the effectiveness of service provided by various service stations.

H_1 : There is difference amongst service station with regard to effectiveness of service. Since the calculated F = 9.95 exceeds table F (9, 36) = 2.15 at 5% level, reject H_0 and accept H_1.

Interpretation : With a confidence level of 95%, it can be said that there is difference in the effectiveness of service provided by the various service stations.

Block (Row) variation :

H_0 : There is no difference between rating of service effectiveness by the 5 professionals.

H_1 : There is difference between rating by the 5 professionals.

Since calculated F = 1.31 does not exceed the table F (4, 36) = 2.36 at 5% level, we have no evidence to reject H_0.

Interpretation : With a confidence level of 95% we conclude that there is no difference in rating by the 5 professionals with regard to effectiveness of service.

12.3 REGRESSION ANALYSIS

Regression analysis is probably the most widely applied technique amongst the analytical models of association used in marketing research. Regression analysis attempts to study the relationship between a dependent variable and a set of independent variables (one or more). For example, in demand analysis, demand is inversely related to price for normal commodities. We may write $D = A - BP$ where D is the demand which is the dependent variable, P is the unit price of the commodity, an independent variable. This is an example of a simple linear regression equation. The multiple linear regressions model is the prototype of single criterion/multiple predictor association model where we would like to study the combined influence of several independent variables upon one dependent variable. In the above example if P is the consumer price index, and q is the index of industrial production, we may be able to study demand as a function of two independent variables P and Q and write $D = A - BP + C Q$ as a multiple linear regression model.

The objectives of the market researchers in using Regression Analysis are :

1) To study a general underlying pattern connecting the dependent variable and independent variables by establishing a functional relationship between the two In this equation the degree of relationship is derived which is a matter of interest to the researcher in his study.

2) To use the well established regression equation for problems involving prediction and forecasting.

3) To study how much of the variation in the dependent variable is explained by the set of independent variables. This would enable him to remove certain unwanted variables from the system. For example if 95% of variation in demand in a study could be explained by price and consumer rating index, the researcher may drop other factors like industrial production, extent of imports, substitution effect etc. which may contribute only 5% of variation in demand provided all the causal variables are linearly independent.

We proceed by first discussing bivariate (simple) regression involving the dependent variables as a function of one independent variable and then on to multiple regression.

Simple linear regression model is given by

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

where Y is the dependent variable,

X_1 is independent variable

ϵ is a random error term

β_0 and β_1 are the regression coefficients to be estimated.

Assumptions of the model :

1) The relationship between Y and X_1 is linear.

2) Y is a random variable which follows a normal distribution from which sample values are drawn independently.

3) X_1 is fixed and is non-stochastic (non-random).

4) The means of all these normal distribution of Y as conditioned by X_1, lie on a straight line with slope β_1.

5) ϵ is the error term $\cap$ IND $(0, \sigma^2)$ and independent of X_1.

Computational aspect :

Estimated regression line based on sampling is written as $\hat{Y} = a + bX_1$

a and b are estimates of β_0 and β_1 obtained through the method of least square by minimising the error sum of squares.

We state the normal equations without going into any derivations. The normal equations are

$$\sum Y = na + b \sum X_1$$

$$\sum X_1 Y = a \sum X_1 + b \sum X_1^2$$

Solve these two simultaneous equations you get the values of a and b.

Total sum of squares = $\Sigma (Y - \bar{Y})^2$
(TSS)

Regression sum of squares = $\Sigma (\hat{Y} - \bar{Y})^2$
(RSS)

Error sum of squares = $\Sigma (Y - \hat{Y})^2$
(ESS)

Form the ANOVA Table for Regression

Source	D.F.	Sum of squares	Mean squares	F ratio
Due to Regression	1	RSS	$\frac{RSS}{1}$	$\frac{RSS}{1} \div \frac{ESS}{(n-2)}$
Due to Error	n–2	ESS	$\frac{ESS}{n-2}$	
Total	n–1	TSS		

$H_0 \beta_1 = 0$ There is no linear relationship between Y and X_1 (Y & X_1 are independent).
$H_1 \beta_1 \neq 0$ There is linear relationship between Y and X_1 as stated in our model.

If the calculated F exceeds Table F (1, n – 2) at 5% level reject H_0 and accept H_1.

Strength of association

It is one thing to find the regression equation after validating the linearity relationship; but at this point we still do not know how strong the association is. In other words, how well does X_1 predict Y?

This is measured by the co-efficient of determination

$r^2 = \frac{RSS}{TSS}$ = Variation in Y explained by regression compared to total variation.

Higher the r^2, greater is the degree of relationship.

The product moment correlation or simple correlation co-efficient between Y and X_1

is = $\sqrt{r_2} = \sqrt{\frac{RSS}{TSS}}$

r^2 lies between 0 and 1. 0 measuring no correlation and 1 measuring perfect correlation.

r lies between –1 and +1 and the sign of r is determined by the sign of the sample regression coefficient (b) in the sample regression equation

$\hat{Y} = a + bX_1$

Having given a foundation structure with underlying assumptions and possible analysis of the model, we now turn our attention to a numerical example to clarify the concepts. It is needless to mention that analysis of data and interpretation of the results are of paramount importance.

Suppose that a marketing researcher is interested in consumers attitude towards nutritional diet of a ready to eat cereal.

X_1 : the amount of protein per standard serving

In the nature of a pretest, the researcher obtains consumer's interval-scaled evaluation of the ten concept descriptions, on a preference rating scale ranging from 1, dislike extremely, upto 9, like extremely well. The data is given below.

Rater	Preference rating (Y)	Protein X_1	
1	3	4	$\Sigma Y = 43\ \bar{y} = 4.3$
2	7	9	
3	2	3	$\Sigma X_1 = 43\ \bar{x}_1 = 4.3$
4	1	1	
5	6	3	$\Sigma YX_1 = 247$
6	2	4	
7	8	7	$\Sigma X_1^2 = 255$
8	3	3	
9	9	8	
10	2	1	

i) Fit a linear regression model of Y on X_1.

ii) Test the validity of the equation statistically.

iii) What do you think of the strength of association?

Answer

i) The normal equations are

$\Sigma Y = na + b\Sigma x_1$ Here n = 10

$\Sigma X_1 Y = a\ \Sigma X_1 + b\ \Sigma X_1^2$

i.e. $10a + 43b = 43$

$43a + 255b = 247$

solving these two simultaneous equations
we have b = 0.886
a = 0.491

Regression Equation is $\hat{Y} = 0.491 + 0.886X_1$
substitute for all x_i^s to get y^s

The regression coefficient b = 0.886 indicates the change in Y per unit change in X_1

ii) Validity of the equation
ANOVA calculation

Total sum of squares (TSS) = $\Sigma (Y - \bar{Y})^2$
$= (3 - 4.3)^2 + (7 - 4.3)^2 + \ldots\ldots + (2 - 4.3)^2$
$= 76.10$

*Regression sum of squares (RSS) = $\Sigma (Y - \bar{Y})^2$
$= (4.034 - 4.3)^2 + (8.464 - 4.3)^2 + \ldots\ldots$
$+ (1.377 - 4.3)^2 = 55.01$

* Error sum of squares = $\Sigma (\hat{Y} - Y)^2 = (-1.034)^2 + (-1.464)^2 + \ldots\ldots$
$+ (0.623)^2 = 21.09$

*Table of Actual Vs predicted

Actual Y	Predicted Y	Error Y − Ŷ
3	4.034	−1.034
7	8.464	−1.464
2	3.148	−1.148
1	1.377	−0.377
6	3.148	2.852
2	4.034	−2.034
8	6.692	1.308
3	3.148	−0.148

9	7.579	1.422
2	1.377	0.623

$\bar{y} = 4.3$

ANOVA Table

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F ratio
Due to regression	1	55.01	$\frac{55.01}{1} = 55.01$	$\frac{55.01}{2.64} = 20.84$
Due to error	8	21.09	$\frac{21.09}{8} = 2.64$	
Total	9	76.10		

H_0 : Where is no relationship of linear type between Y and X_i
i.e. $\beta_i = 0$
H_1 : Y is linearly related to X_1 $\beta \neq 0$

Interpretation of results
Calculated F = 20.84 exceeds Table F (1,8) = 5.32 at 5 % level. Reject H_0 and accept H_1. We conclude Y is a linear function of X_1 with a confidence level of 95%. In other words preference rating Y is linearly related to amount of (X_1) Protein per standard serving of the cereal with a confidence level of 95%. Thus the equation is a valid one.

iii) Strength of association

Coefficient of determination $r^2 = \frac{RSS}{TSS} = \frac{55.01}{76.10} = 0.723$

This implies that 72.3% of the variation in Y is explained by the regression and only 27.7% of the variation is explained by error. The association is strong to enable X_1 to predict Y.

$r = \sqrt{r^2} = 0.850$

Here r will have a positive sign since b is positive. Before starting our discussion on the multiple regression, let us give a brief account of the usefulness of simple linear regression in sales forecasting using time series data in which time t is an independent variable.

1) Linear Trend
$Y = a + bt$

If Y represents the sales data collected for the past many years for example last 10 years from 1979 to 1988, we normalise the year by taking t = 1 corresponding to 1979, t = 2 for 1980 etc. and t = 10 for 1988. Now the simple linear regression model can be directly applied to forecast Y (sales) say for 1989 after fitting the regression equation.

2) Trend in semilog form
$Y = ab^t$
Taking log on both sides we have
Log Y = Log a + t log b
This reduces to $Z = A + Bt$
where $Z = \text{Log } y$
$A = \text{Log } a$
$B = \text{Log } b$

This can now be solved as a simple linear regression model for forecast where Z is dependent variable and t is independent variable as before.

3) Double log form
$Y = at^b$
Log Y = log a + b log t
i.e. $Z = A + bT$
where Z = log Y A = log a_i T = Log t.

This can now be solved as normal bivariate regression equation to forecast sales for the next period.

It is now time to introduce the concept of multiple regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_k X_k + \epsilon$$

The assumptions are exactly same as simple linear regression except that you add X_1, X_2,.....X_k in the place of X_1 because Y is linearly related to X_1.....X_k and our aim is to understand the combined influence of the K factors X_1, X_2..... X_k on Y. To understand the concept clearly, let us study a case of 2 independent variables and write the model as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

so that $\hat{Y} = a + b X_1 + cX_2$ being the estimated regression equation where we add one more independent variable X_2 in the model. Suppose we extend the previous example of bivariate regression on preference rating Vs. protein (X_1) by adding X_2 the percentage of minimum daily requirements of vitamin D per standard serving. Let us see how the multiple regression model emerges to explain the variation in the dependent variable Y caused by X_1 and X_2. Let us look at the following table giving the data on Y, X_1, and X_2.

Rater	Preference Rating Y	Protein X_1	Vitamin D X_2	
1	3	4	2	
2	7	9	7	$\Sigma Y = 43$
3	2	3	1	$\Sigma X_1 = 43$
4	1	1	2	$\Sigma X_2 = 40$
5	6	3	3	
6	2	4	4	$\Sigma YX_1 = 247$
7	8	7	9	$\Sigma YX_2 = 232$
8	3	3	2	$\Sigma X_1^2 = 255$
9	9	8	7	$\Sigma X_2^2 = 226$
10	2	1	3	$\Sigma X_1X_2 = 229$

The normal equations are :

$\Sigma Y = Na + b \Sigma X_1 + c \Sigma X_2$

$\Sigma YX_1 = a \Sigma X_1 + b \Sigma X_1^2 + c \Sigma X_1X_2$

$\Sigma YX_2 = a\Sigma X_2 + b \Sigma X_1X_2 + c \Sigma X_2^2$

$10a + 43b + 40c = 43$

$43a + 255b + 229c = 247$

$40a + 229b + 226c = 232$

Solving for a, b and c we have

$a = 0.247$

$b = 0.493$

$c = 0.484$

$\hat{Y} = 0.247 + 0.493 X_1 + 0.484 X_2$

Here b and c are called partial regression coefficients b = 0.493 denotes the change in $\hat{Y}$ per unit change in X_1 when X_2 is held constant. Similarly c = 0.484 denotes the change in $\hat{Y}$ per unit change in X_2 when X_1 is held constant.

By now you must have notified the cumbersome calculations involved when the number of variables increase and becomes extremely difficult when the number of variables is more than 3. One has to resort to computer based regression models. In fact it may be mentioned here that all multivariate procedures require the help of computer when the variables and observations are large. As before we can calculate the other co-efficients like R^2 co-efficient of multiple determination and R multiple correlation co-efficient and also ANOVA for hypothesis testing. The author has developed his own user friendly program for multiple regression with a conversational

style based on IBM PC MS DOS. We will use the output of the program and interpret the results of our problem which is the most important aspect for us.

Multiple Linear Regression

Number of Variables ? 3
Number of Observations ? 10

WANT TO CHANGE NUMBER OF VARIABLES/NUMBER OF OBSERVATIONS (Y OR N) ? N
NAME OF VARIABLE # 1 ? Y
NAME OF VARIABLE # 2 ? X1
NAME OF VARIABLE # 3 ? X2

DATA GATHERED FOR VARIABLE Y :
— PERIOD # 1 ? 3
— PERIOD # 2 ? 7
— PERIOD # 3 ? 2
— PERIOD # 4 ? 1
— PERIOD # 5 ? 6
— PERIOD # 6 ? 2
— PERIOD # 7 ? 8
— PERIOD # 8 ? 3
— PERIOD # 9 ? 9
— PERIOD # 10 ? 2

DATA GATHERED FOR VARIABLE X1 :
— PERIOD # 1 ? 4
— PERIOD # 2 ? 9
— PERIOD # 3 ? 3
— PERIOD # 4 ? 1
— PERIOD # 5 ? 3
— PERIOD # 6 ? 4
— PERIOD # 7 ? 7
— PERIOD # 8 ? 3
— PERIOD # 9 ? 8
— PERIOD # 10 ? 1

DATA GATHERED FOR VARIABLE X2 :
— PERIOD # 1 ? 2
— PERIOD # 2 ? 7
— PERIOD # 3 ? 1
— PERIOD # 4 ? 2
— PERIOD # 5 ? 3
— PERIOD # 6 ? 4
— PERIOD # 7 ? 9
— PERIOD # 8 ? 2
— PERIOD # 9 ? 7
— PERIOD # 10 ? 3

Correlation Matrix

1	.85	.85
.85	1	.84
.85	.84	1

Variance Covariance Matrix

2.91	6.9	6.67
6.9	2.79	6.33
6.67	6.33	2.71

Variable	Mean	Std Deviation
Y	4.3	2.907844
X1	4.3	2.790858
X2	4	2.708013

CONTINUE ? Y
CONTINUE ? Y

Regression Equation

Dependent Variable : Y

Independent variable	ESTIMATED COEFFICIENT	BETA %	Errors	T-Test
X1	.49	.49	34	1.47
X2	.48	.45	35	1.4
CONSTANT	.2471704			

Determination Coefficient	= .78
Correlation Coefficient	= .89
F-Test	= 12.65
Degrees of Freedom	= 2, 7
Sum of squares of error	= 16.49

CONTINUE?
CONTINUE ? Y

TABLE OF RESIDUAL VALUES

#	Observation	Estimation	Residual
1	3	3.18	- .18
2	7	8.07	-1.07
3	2	2.21	- .21
4	1	1.71	- .71
5	6	3.18	2.82
6	2	4.15	-2.15
7	8	8.05	- .05
8	3	2.69	.31
9	9	7.57	1.43
10	2	2.19	- .19

CONTINUE ? Y

Analysis of Variance Table

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Due to Regression	2	59.61	29.81	12.65
Due to Error	7	16.49	2.36	
Total	9	76.1		

Another Analysis (Type Y or N)?

The program output gives many other statistical analysis which we will not touch upon now, and come to our important tests straightway. The residual or error between Y and $\hat{Y}$ i.e. between actual and forecast on important measure of reliability of the model is printed out for each observation. If you look at the errors, you get a fairly good idea about the model equation. However for validity of the regression equation, you look first at the coefficient of multiple determination R^2 and multiple correlation coefficient R. In our example $R^2 = 0.78$ and $R = 0.89$ which is a satisfactory one indicating that the preference rating Y is linearly related to protein intake X_1 and vitamin D intake X_2. It tells that 78% of variation in Y is explained jointly by the variations in X_1 and X_2 jointly.

Hypothesis testing for linearity through ANOVA.

$H_0 : \beta_1 \& \beta_2 = 0 \Rightarrow$ There is no linear relation between Y, X_1 and X_2.
$H_1 : \beta_1 \& \beta_2 \neq 0 \Rightarrow$ There is linear relationship.

Look at the ANOVA
The calculated F = 12.65
The table F (2, 7) = 4.74 at 5% level. Reject H_0 and accept H_1.

That is the linear regression equation between Y & X_1, X_2 is statistically valid. We are 95% confident that preference rating is linearly related to X_1 and X_2 and the equation is $\hat{Y} = 0.247 + 0.49X_1 + 0.48X_2$

Points to ponder on Multiple regression analysis

1) Equation should be validated statistically.

2) For forecasting the dependent variable, the independent variables should be forecast first.
For example if demand is a function of price index and production index established by a multiple regression model, then to forecast demand for the next period, it is imperative first to forecast the price index and production index and then substitute them in the model to get the forecast for demand. This is one of the limitations of regression forecasting.

3) When the variables become too many the analysis is complex and very often the market researcher does not know which variables to retain. This problem could be overcome by doing 'stepwise regression' on computer. For example if demand is a function of 20 variables, we first fit demand equation with 3 important variables which we think affect demand. Suppose $R^2 = 0.85$ that is 85% of the variation in demand is explained by these variables, we add another two more variables of importance to make five independent variables. Now if $R^2 = 0.95$ we can as well stop adding further variables as the contribution may not appreciably improve the situation. We can thus visualise demand as a function of just 5 variables. The various permutations of changing and adding variables is possible only with the help of a computer. The important point to remember is that the cut off point for the number of variables to be added should be based on the increase every time you get on R^2. The moment the increase is marginal, stop adding variables.

4) If the independent variables among themselves are highly correlated, then we are facing the problem of 'multicollinearity'. Normally we say that the partial regression coefficient with respect to X_1 implies change in Y per unit change in X_1 provided X_2, X_3..... are held constant. This poses a serious problem if there is multicollinearity. One way to overcome multicollinearity is to drop certain variables from the model if the corresponding standard error of regression coefficient are unduly large. Another method is to see whether the original set could be transformed into another linear composite so that the new variables are uncorrelated.

12.4 DISCRIMINANT ANALYSIS

It has been pointed out earlier, that the discriminant analysis is a useful tool for situations where the total sample is to be divided into two or more mutually exclusive and collectively exhaustive groups on the basis of a set of predictor variables. For example, a problem involving classifying sales people into successful and unsuccessful; classifying customers into owners or and non-owners of video tape recorder, are examples of discriminant analysis.

Objectives of two group discriminant analysis :

1 Finding linear composites of the predictor variables that enable the analyst to separate the groups by maximising among groups relative to with in-groups variation.

2 Establishing procedures for assigning new individuals, whose profiles but not group identity are known, to one of the two groups.

3 Testing whether significant differences exist between the mean predictor variable profiles of the two groups.

4 Determining which variables account most for intergroup differences in mean profiles.

A numerical example

Let us return to the example involving ready-to-eat cereal that was presented in the regression analysis. However in this problem the ten consumer raters are simply asked to classify the cereal into one of two categories like versus dislike. The data is given below : Here again

X_1 : The amount of protein (in grams) per standard serving.
X_2 : The percentage of minimum daily requirements of vitamin D per standard serving.

Also shown in the data table are the various sums of squares and cross products; the means on X_1 and X_2 of each group, and total sample mean.

Consumer evaluations (like versus dislike) of ten cereals varying in nutritional content

Person	Evaluation	Protein X_1	Vitamin D X_2	X_1^2	X_2^2	X_1X_2
1	Dislike	2	4	4	16	8
2	Dislike	3	2	9	4	6
3	Dislike	4	5	16	25	20
4	Dislike	5	4	25	16	20
5	Dislike	6	7	36	49	42
	Mean	4	4.4	Sum 90	110	96
6	Like	7	6	49	36	42
7	Like	8	4	64	16	32
8	Like	9	7	81	49	63
9	Like	10	6	100	36	60
10	Like	11	9	121	81	99
	Mean	9	6.4 Sum	415	218	296
	Grand Mean	6.5	5.4			
	Standard deviation	3.028	2.011			

The grand mean is $\bar{X}_1 = 6.5$, $\bar{X}_2 = 5.4$

We first note from the table that the two groups are much more widely separated on X_1 (protein) than they are on X_2 (Vitamin D). If we were forced to choose just one of the variables, it would appear that X_1 is a better bet than X_2. However there is information provided by the group separation on X_2, so we wonder if some linear composite of both X_1 and X_2 could do better than X_1 alone. Accordingly we have the following linear function :

$Z = K_1X_1 + K_2X_2$ where K_1 and K_2 are the weights that we seek.

But how shall we define variability? In discriminant analysis, we are concerned with the ratio of two sums of squares after the set of scores on the linear composite has been computed. One sum of squared deviations represents the variability of the two group means on the composite around their grand mean. The second sum of squared deviations represents the pooled variability of the individual cases around their respective group means also on the linear composite. One can then find the ratio of the first sum of squares to the second. It is this ratio that is to be maximised through the appropriate selection of K_1 and K_2. Solving for K_1 and K_2 involves a procedure similar to the one encountered in the multiple regression. However in the present case, we shall want to find a set of sums of squares and cross products that relate to the variation within groups. For ease of calculation let us define $x_1 = X_1 - \bar{X}_1$ and $x_2 = X_2 - \bar{X}_2$ (i.e. each observation measured from its mean)

Solving for K_1 and K_2

Mean corrected sums of squares and cross products.

	Dislikers	Likers	Total
$\Sigma x_1^2 = \Sigma (X_1 - \bar{X}_1)^2 = \Sigma X_1^2 - N\bar{X}_1^2 =$	10	10	20
$\Sigma x_2^2 = \Sigma (X_2 - \bar{X}_2)^2 = \Sigma X_2^2 - N\bar{X}_2^2 =$	13.2	13.2	26.4
$\Sigma x_1x_2 = \Sigma (X_1 - \bar{X}_1)(X_2 - \bar{X}_2) = \Sigma X_1X_2 - N\bar{X}_1\bar{X}_2 =$	8	8	16

The normal equations are

$K_1 \Sigma x_1^2 + K_2 \Sigma x_1 x_2 = \bar{X}_1$ (Likers) $- \bar{X}_1$ (dislikers)
$K_1 \Sigma x_1 x_2 + K_2 \Sigma x_2^2 = \bar{X}_2$ (Likers) $- \bar{X}_2$ (dislikers)
$20 K_1 + 16 K_2 = 9 - 4 = 5$
$16 K_1 + 26.4 K_2 = 6.4 - 4.4 = 2$

Solving these two simultaneous equations, we have $K_1 = 0.368, K_2 = -0.147$
Discriminant function $Z = 0.368X_1 - 0.147X_2$

We can also find discriminants scores for the means of the two groups and the grand mean.
$\bar{Z}$ (dislikers) $= 0.368\ (4) - 0.147\ (4.4) = 0.824$
$\bar{Z}$ (likers) $= 0.368\ (9) - 0.147\ (6.4) = 2.368$
$\bar{Z}$ (grand means) $= 0.368\ (6.5) - 0.147\ (5.4) = 1.596$

We note that the discriminant function "favours" X_1 by giving about 2.5 times the (absolute value) weight ($K_1 = 0.368$ versus $K_2 = -0.147$) to X_1 as is given to X_2.

The discriminant scores of each person are shown below. Each score is computed by the application of the discriminant function to the persons original X_1 and X_2 values.

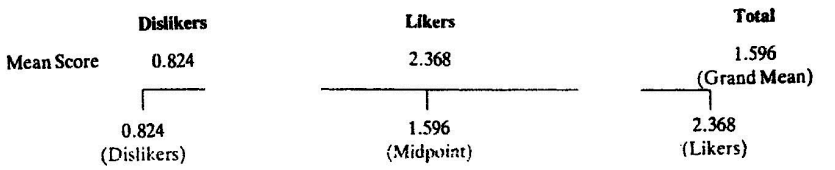
Dislikers		Likers	
Person	**Discriminant Score**	**Person**	**Discriminant Score**
1	0.148	6	1.691
2	0.809	7	2.353
3	0.735	8	2.279
4	1.250	9	2.794
5	1.176	10	2.721
Mean	0.824	Mean	2.368
	Grand Mean 1.596		

Between group variability :
$5\ (0.824 - 1.596)^2 + 5\ (2.368 - 1.596)^2 = 5.96$
Within group variability
Dislikers $(0.148 - 0.824)^2 + (0.809 - 0.824)^2 + \ldots\ldots (1.176 - 0.824)^2 = 0.772$
Likers $(1.691 - 2.368)^2 + (2.353 - 2.368)^2 + (2.721 + 2.368)^2$ $\quad \frac{0.772}{1.544}$

Discriminant criterion $C = \frac{5.96}{1.544} = 3.86$

Since the normal equations for solving k_1 and k_2 are obtained by maximising the ratio between group and within group variance the discriminant criterion as calculated above = 3.86 will be the maximum possible ratio. If we suppress X_2 in the discriminant function and calculate another C, it will be less than 3.86. It is rather interesting that the optimal function $Z = 0.368\ X_1 - 0.147\ X_2$ is a difference function in which X_2 (Vitamin D) receives a negative weight bringing thereby the importance of X_1 to the highest order. This means protein is much more important than Vitamin D.

Classifying the persons
It is all well and good to find the discriminant function, but the question is how to assign the persons to the relevant groups.



- Assign all cases with discriminant scores that are on the left of the midpoint (1.596) to the disliker group.

- Assign all cases with discriminant scores that are on the right of the midpoint (1.596) to the liker group.

That is all true dislikers will be correctly classified as such and all true likers will be correctly classified. This can be shown by a 2×2 table :

	Assigned by Rule		
True State	Disliker	Liker	Total
Disliker	5	0	5
Liker	0	5	5
Total	5	5	10

Testing Statistical Significance

While the discriminant function does perfectly in classifying the ten cases of the illustration on protein (X_1) and vitamin (X_2) into likers and dislikers, we still have not tested whether the group means differ significantly. This is also based on F ratio which required calculation of Mahalanobis D^2. This calculation of F is a little complicated which is normally an output parameter in the standard package like Biomedical computer program and SPSS of IBM. Biomedical computer program of the University of California press is an outstanding software containing all multivariate procedures. For our illustration let us calculate F

$$F = \frac{n_1 n_2 (n_1 + n_2 - m - 1)}{m (n_1 + n_2)(n_1 + n_2 - 2)} \cdot D^2$$

$$\sim F \text{ distribution with } m_1\ n_1 + n_2 - m - 1 \text{ d.f.}$$

where n_1 = number of observations in group 1
n_2 = number of observations in group 2
m = number of independent variables
D^2 = Mahalanobis square distance

In our problem $n_1 = 5$
$n_2 = 5$
$m = 2$ (X_1 and X_2)

Simple way of calculating D^2 would be to use the discriminant function

$$D^2 = (n_1 + n_2 - 2)\,(0.368\,(5.0) - 0.147\,(2))$$
$$= 8\,(0.368 \times 5 - 0.147 \times 2) = 12.353$$

You please note that the expression within brackets is the discriminant function $Z = 0.368\,X_1 - 0.147X_2$ where X_1 and X_2 are substituted by the respective group means difference : $\bar{X}_1$ (likers) – $\bar{X}_1$ (dislikers), $\bar{X}_2$ (likers) – $\bar{X}_2$ (dislikers)

$$\therefore F = \frac{5 \times 5\,(5+5-2-1)}{2 \times (5+5)\,(5+5-2)} \times 12.353$$

$$= \frac{25 \times 7}{2 \times 10 \times 8} \times 12.353 = 13.511$$

Table F (2, 7) = 4.74 at 5% level.

Since the calculated F exceeds table F at 5% level, reject H_0 and accept H_1 i.e. the group means are not equal in importance with a probability of 95%. This clearly validates the relative importance of X_1 far higher than X_2.

Brief remarks on Multiple Discriminant Analysis :

You would have realised by now the complexity of calculations in the discriminant analysis involving 2 predictors which itself needs computer based solutions when the number of observations increases considerably. Multiple discriminant an analysis is invariably carried out by means of computer programs. One of the most flexible and comprehensive programs in BMD-07M of the biomedical program series of the

University of California press. SPSS also has all multivariate procedures. It may be mentioned that the basic structure of the bivariate analysis remains same in multiple case also. What is important for you is interpretation of the results and findings of the study.

Activity 2

What are the differences between Regression Analysis and Discriminant Analysis?

.....

.....

.....

.....

.....

.....

.....

12.5 FACTOR ANALYSIS

Factor analysis is a generic name given to a class of techniques whose purpose is data reduction and summarisation. Very often market researchers are overwhelmed by the plethora of data. Factor analysis comes to their rescue in reducing the number of variables. Factor analysis does not entail partitioning the data matrix into criterion and predictor subsets; rather interest is centred on relationships involving the whole set of variables. In factor analysis :

1. The analyst is interested in examining the "strength" of the overall association among variables in the sense that he would like to account for this association in terms of a smaller set of linear composites of the original variables that preserve most of the information in the full data set. Often his interest will emphasize description of the data rather than statistical inference.
2. No attempt is made to divide the variables into criterion versus prediction sets.
3. The models are primarily based on linear relationships.

Factor analysis is a "search" technique. The researcher-decision maker does not typically have a clear a priori structure of the number of factors to be identified. Cut off points with respect to stopping rules for the analysis are often ad hoc as the output becomes available. Even where the procedures and rules are stipulated in advance, the results are more descriptive than inferential.

The procedure involved in computation of factor analysis is extremely complicated and cannot be carried out effectively without the help of computer. Packages like SPSS, SAS and Biomedical programs (BMD) can be used to analyse various combinations leading to factor reduction. We will make an attempt to conceptualise the scenario of factor analysis with emphasis on the interpretation of figures.

The term "factor analysis" embraces a variety of techniques. Our discussion focuses on one procedure : principal component analysis and the factors derived from the analysis are expressed as linear equations. These linear equations are of the form

$$F_i = a_{1i}\, x_1 + a_{2i}\, x_2 + a_{3i}\, x_3 + \ldots\ldots + a_{mi} x_m$$

The i factors are derived, and each variable appears in each equation. The a-coefficients indicate the importance of each variable with respect to a particular factor coefficient of zero indicating the variable is of no significance for that factor. In principal component analysis, the factors are derived sequentially, using criteria of maximum reduction in variance and non-correlation among factors.

Let us go to a specific example to explain factor analysis and its output.

Example

A manufacturer of fabricating parts is interested in identifying the determinants of a successful salesperson. The manufacturer has on file the information shown in the following table. He is wondering whether he could reduce these seven variables to two or three factors, for a meaningful appreciation of the problem.

Data Matrix for Factor Analysis of seven variables (14 sales people)

Sales person	Height (x_1)	Weight (x_2)	Education (x_3)	Age (x_4)	No. of Children (x_5)	Size of household (x_6)	IQ (x_7)
1	67	155	12	27	0	2	102
2	69	175	11	35	3	6	92
3	71	170	14	32	1	3	111
4	70	160	16	25	0	1	115
5	72	180	12	36	2	4	108
6	69	170	11	41	3	5	90
7	74	195	13	30	1	2	114
8	68	160	16	32	1	3	118
9	70	175	12	45	4	6	121
10	71	180	13	24	0	2	92
11	66	145	10	39	2	4	100
12	75	210	16	26	0	1	109
13	70	160	12	31	0	3	102
14	71	175	13	43	3	5	112

Can we now collapse the seven variables into three factors? Intuition might suggest the presence of three primary factors : A maturity factor revealed in age/children/size of household, physical size as shown by height and weight, and intelligence or training as revealed by education and IQ.

The sales people data have been analysed by the SAS program. This program accepts data in the original units, automatically transforming them into standard scores. The three factors derived from the sales people data by a principal component analysis (SAS program) are presented below :

Three-factor results with seven variables.

Sales people characteristics

Variable	Factor I	II	III	Communality
Height	0.59038	0.72170	−0.30331	0.96140
Weight	0.45256	0.75932	−0.44273	0.97738
Education	0.80252	0.18513	0.42631	0.86006
Age	−0.86689	0.41116	0.18733	0.95564
No. of children	−0.84930	0.49247	0.05883	0.96730
Size of household	−0.92582	0.30007	−0.01953	0.94756
IQ	0.28761	0.46696	0.80524	0.94918
Sum of squares	3.61007	1.85136	1.15709	
Variance summarised	0.51572	0.26448	0.16530	0.94550

Factor Loadings : The coefficients in the factor equations are called "factor loadings" They appear above in each factor column, corresponding to each variable. The equations are :

$$F_1 = 0.59038x_1 + 0.45256x_2 + 0.80252x_3 - 0.86689x_4 - 0.84930x_5 - 0.92582x_6 + 0.28761x_7$$

$$F_2 = 0.72170x_1 + 0.75932x_2 + 0.18513x_3 + 0.41116x_4 + 0.49247x_5 + 0.30007x_6 + 0.46696x_7$$

$$F_3 = -0.30331x_1 - 0.44273x_2 + 0.42631x_3 + 0.18733x_4 + 0.5883x_5 - 0.01953x_6 + 0.80524x_7$$

The factor loadings depict the relative importance of each variable with respect to a particular factor. In all the three equations, education (x_3) and IQ (x_7) have got positive loading factor indicating that they are variables of importance in determining the success of sales person.

Variance summarised : Factor analysis employs the criterion of maximum reduction of variance — variance found in the initial set of variables. Each factor contributes to reduction. In our example Factor I accounts for 51.6 per cent of the total variance. Factor II for 26.4 per cent and Factor III, for 16.5 per cent. Together the three factors "explain" almost 95 per cent of the variance.

Communality : In the ideal solution the factors derived will explain 100 per cent of the variance in each of the original variables, "Communality" measures the percentage of the variance in the original variables that is captured by the combination of factors in the solution. Thus a communality is computed for each of the original variables. Each variables communality might be thought of as showing the extent to which it is revealed by the system of factors. In our example the communality is over 85 per cent for every variable. Thus the three factors seem to capture the underlying dimensions involved in these variables.

There is yet another analysis called varimax rotation, after we get the initial results. This could be employed if needed by the analyst. We do not intend to dwell on this and those who want to go into this aspect can use SAS program for varimax rotation.

In the concluding remarks, it should be mentioned that there are two important subjective issues which should be properly resolved before employing factor analysis model. They are :

1) How many factors should be employed in attempting to reduce the data? What criteria should be used in establishing that number?

2) The labelling of the factors is purely intuitive and subjective.

Activity 3

12.6 SUMMARY

We have started our discussion by giving an overview of the various multivariate analysis procedures in the context of associative data with a marketing orientation. We have given a brief introduction of the multivariate tools and their applicability in the relevant problem areas.

We have discussed the concept of analysis of variance. We have clearly brought out the assumptions underlying the one way and two-way classification models and the methodology of separation of total variance into meaningful components variations and error variations. Hypothesis testing using ANOVA table has been clearly explained using examples from marketing which include testing sales and service effectiveness using experimental data.

The next topic of discussion has been the regression analysis, which is probably the most widely used technique amongst the analytical models of association. We have started the simple linear regression model first to introduce the concept of regression and then moved on to the multiple linear regression mode. All the underlying assumptions of the model have been clearly explained. Both the bivariate and multivariate regression models have been illustrated using the example of preference rating as a function of protein intakes, and vitamin D intake perception in the case of a ready to eat cereal. The concept of testing the linear equation, contribution made by regression in explaining variation in dependent variables and strength of association have all been explained using ANOVA table. A brief account of the role of regression in sales forecasting involving time series analysis has also been given. The need for resorting to computer solutions for large number of variables and observations has been brought out with an actual print out of the example already discussed. The concept of stepwise regression and the problems encountered in any regression analysis have also been explained.

Then we have gone to the discriminant analysis technique—a technique when the interest is to classify the groups on the basis of a set of predictor variables. We have explained the concept of separation by giving examples of classifying sales people into successful and unsuccessful, customers into owners and non-owners etc. As before, we have begun the discussion with discriminant function involving two predictor variables using the example of 'ready to eat cereal problem' but with a difference — classifying the persons into liker group and disliker group. The discriminant function, the discriminant criterion and the assignment rule have all been explained testing the statistical significance using F test based on Mahalanobis D^2 has also been carried out. We have pointed out that the multiple discriminant analysis involving more than two predictor variables require the use of computer although the basic structure of the model does not change.

Factor analysis is the last multivariate tool that we have discussed in this unit. We have first mentioned that the fundamental objective of factor analysis is to reduce the number of variables in the data matrix. Then it has been pointed out that the computation of any factor analysis involves very complex calculations which will have to be solved using computer packages like SAS. The concepts of "factor loading", "variance summarised" and "communality" have been clearly explained using one practical example that has been solved by SAS program. The subjective issues like "how many factors?" "what criteria to decide this number?" and "labelling of the factors" have been mentioned at the end.

As concluding remarks, it may be mentioned here that 1) all multivariate procedures can be more effectively solved using standard computer packages when the number of variables and number of observations increase significantly, 2) what is more important is the ability to interpret the results of the market research study involving multivariate analysis.

12.7 SELF-ASSESSMENT EXERCISES

1 a) In a demand forecasting study involving a normal commodity, two simple linear regression models are fitted :

$D = 8.5 + 0.22\ p$ ($r^2 = 0.75$)

$\text{Log } D = 1.3 + 0.10 \log p$ ($r^2 = 0.80$)

i) which model would you prefer and why?

ii) mention the dependent and independent variables.

b) A manufacturer of industrial supplies developed the following model for predicting the number of sales per month

$Y = 41 + .3X_1 + .05X_2 - 7X_3 + 10X_4$

where Y = Sales per month
X_1 = Number of manufacturing firms
X_2 = Number of wholesale and retail firms
X_3 = Number of competing firms
X_4 = Number of full time company sales people.

i) Explain the correct interpretation of all estimated parameters in the equation.
ii) If $R^2 = 0.49$, what does this figure mean to you?
iii) Explain how you will go about testing the validity of this multiple linear regression equation.

2 The following discriminant function was developed to classify sales persons into successful and unsuccessful sales person

$Z = 0.53 X_1 + 2.1X_2 + 1.5X_3$

where X_1 = no. of sales call made by sales persons.
X_2 = no. of customers developed by sales persons.
X_3 = no. of units sold by sales person.

The following decision rule was developed.

if $Z \geq 10$, classify the sales person as successful
if $Z < 10$, classify the sales person as unsuccessful.

The sales persons A and B were considered for promotion on the basis of being classified as successful or unsuccessful. Only the successful sales person would be promoted. The relevant data on A and B is given below :

	A	B
X_1	10	11
X_2	2	1.5
X_3	1	0.5

whom will you promote?

3 A large sample of people were asked to rate how much they liked each of 5 beverages—coffee, tea, milk, fruit juice and soft drinks. Through factor analysis of factor loadings was obtained.

	Factors			
	I	II	III	Communality
Coffee	–.219	.363	– .338	0.2939
Tea	–.137	.682	.307	0.5781
Milk	.514	–.213	–.277	0.3611
Fruit Juice	.485	–.117	115	0.2621
Soft drinks	–.358	–.635	.534	0.8165
Sum of squares	0.6943	1.0592	0.5584	
Variance summarised	0.1389	0.2118	0.1117	0.4624

a) Write the linear equations for all the three factors.

b) Interpret the loading co-efficients, variance summarised and communality values of this table.

12.8 FURTHER READINGS

F.E. Brown, *Marketing Research* Addison-Wesley Publishing Company.
Paul E. Green and Donald S. Tull, *Research for Marketing Decisions* Prentice-Hall of India Pvt. Ltd.
Boyd, Westfall and Stasch, *Marketing Research Text and Cases* D. Irwin Inc. Homewood. Illinios.
U.K. Srivastava, G.V., *Quantitative Techniques for Managerial Decision Making* Wiley Eastern Limited.

UNIT 13 CONJOINT ANALYSIS

Objectives

After going through this unit you should be able to :

- Understand the concept and need of conjoint analysis
- Explain the steps involved in conjoint analysis
- Apply conjoint analysis in different marketing needs.

Structure

13.1 Introduction
13.2 What is Conjoint Analysis
13.3 Steps in the Application of Conjoint Analysis
13.4 Summary
13.5 Self-Assesment Exercises
13.6 Further Readings

13.1 INTRODUCTION

In marketing the need to understand consumer behaviour is paramount. Although consumer behaviour has to be understood in its many facets, the need to discover the attributes that the consumer considers important is quite common. After understanding these attributes the marketers attempt to understand consumers perceptions for the products or brands along these important attributes.

Many research and analytical techniques are available and often employed for these purposes. For discovering the salient attributes, one can ask the consumers directly or use analytical, techniques like Factor Analysis or Multi Dimensional Scaling (MDS). Simple techniques of Depth Interviewing or Focus Group Discussions are found to be quite effective for discovering these dimensions or attributes. Later, the scores of different products along these dimensions are obtained from the consumers directly on Likert type metric scales.

Factor Analysis (discussed in previous unit) is found to be quite powerful in analytically discovering the important attributes through a set of indirect and large number of questions. MDS is also found quite effective in similar situations. In fact, it is particularly effective when it is not considered desirable to probe consumers directly about the attributes. This may be true when breaking the products into attributes and seeking responses along them may be too artificial for the consumers. A more realistic way may be to ask the consumers to respond to the totality of the products or brands – something that they always do in real life purchase situations.

MDS (which would be discussed in subsequent unit in detail) is found to be superior to more traditional method of attributes wise data collection method because of its lesser demands on data scaling. It can work with metric as well as ranking data. This feature is also very consumer friendly as the consumers, in most cases, compare products in terms of "better" or "worse" instead of assigning absolute values to products or its attributes. For example, a shirt customer may consider the brand, material of the fabric, design, colour, texture of the cloth, price and stitching as important. He shall consider all these attributes when he chooses the particular shirt for his purchase out of the set of shirts displayed to him in the shop. But, he many find it difficult to accurately provide his ratings for each of these attributes for each of the shirt displayed to him.With the help of above techniques, we can understand the significant attributes of products and obtain the plots of competing products or brands in the perceptual space of consumers. These information go a long way in helping us to take the product positioning decisions.

While it is very useful for the makerter to see the plots of existing products and visualise alternative positions together with their respective sizes, it is not sufficient for taking some of the important managerial decisions. For example, it does not help him in understanding

the trade-offs that the consumers make between different attributes. In fact. it is these trade-offs with which the marketer find his greatest opportunities to optimise his marketing efforts. Sometimes, he may not even find it feasible to exploit the precise opportunity that is suggested by the methods like MDS or direct plotting of consumers perceptual space. For example, the available technology to produce the product may allow some variant of the product instead of the ones thrown up by the traditional methods.

The marketer is also naturally interested in getting the answers to several "What if" kind of questions. What if I reduce one attribute? What if the competitor changes its brand position? What if the salience of one attribute changed through some consumer education efforts? Etc.

Conjoint Analysis technique has been developed to answers most of these questions and guide the marketer in optimising his decisions. Yet, the data requirements on the consumers are minimised and anchored at a very natural level. This techniques was introduced to the marketers in 1970s and has gained tremendous popularity since then due to its great managerial relevance and simplicity.

13.2 WHAT IS CONJOINT ANALYSIS

Conjoint Analysis is basically a data decompositional technique which tries to plot the output data on the joint space of the importance of each attribute are the attribute. It seeks data from the consumers in the form of their overall response to the totality of products (or their descriptions) while the output is in terms of the scores that the consumer has implicitly assigned to each of the attribute and its levels.

The important thing to note is that the consumer is not asked to assign scores to different attributes separately. In fact, the consumer is presented the stimulus in the form of totality of the product like in the case of MDS. However, there is one difference between and Conjoint Analysis. In Conjoint Analysis, the stimuli are created by the researcher himself. While in the case of MDS already existing products or brands are used.

Conjoint Analysis derives the importance weights (called "part worth utilities") assigned by each consumer to respective levels of attributes in such a way that they are directly comparable. This feature of the technique allows to determine the trade-offs that the consumers make in their minds. The relative importance of the attributes can also be derived from the output of "part worth utilities" Thus, starting with a very simple input data (just the ranking of some predesigned product alternatives) Conjoint Analysis provides the part worth utilities for each of the product attributes levels for every consumer individually.

13.3 STEPS IN THE APPLICATION OF CONJOINT ANALYSIS

The main steps involved in the application of Conjoint Analysis are following:

1. Determination of the salient attributes for the given product from the points of view of the consumers
2. Assigning a set of discrete levels or a range of continuous values to each of the attributes.
3. Utilising Fractional Factorial Design of Experiment for designing the stimuli for experiment.
4. Physically designing the stimuli
5. Ranking or rating data collection.
6. Conjoint analysis and determination of part worth utilities.
7. Applying conjoint analysis output for different marketing decisions.

We shall now describe each of these steps in some details.

1. Identification of Salient Attributes

Unlike MDS or Factor Analysis, Conjoint Analysis requires the salient attributes to be given. These attributes can be selected based on the marketer's experience, depth interviewing or Focussed group discussions. In some cases Factor Analysis or MDS may also be used for labelling the salient factors. Only those attributes should be selected here around which differences do exist among brands or through which future differentiations can be achieved. For, if it is not possible to differentiate products along any particular attribute, the managerial value of that attribute becomes very low.

2. Assignment of levels to selected attributes

The real products can be described as specific combinations of the attributes where each attribute can take different values. Sometimes the attribute can take only discrete values. Examples of such attributes may be brand name , colour of the product or the nature of technology used. Out of the entire range of values that may be theoretically possible, the marketer may choose only a few for his active consideration. For example, a soap can technically be produced in all colours. But, some colours, like dark black, may be ruled out.

Some attributes, like price or life of the product, may take continuous values. Conjoint Analysis can treat them or like that also. Some range of values may be specified for them or the range may be kept open ended.

3. Fractional Factorial Design of Experiment

In Conjoint Analysis the profile of different products are presented to the consumers for their responses. These profiles are generated by varying the levels of its attributes. For example, suppose we are conducting a Conjoint Analysis based study of dish washers. Let us assume that the most important attributes considered by its customers are Brand, Price, Washing Capacity, Colour and Shape. Let us further assume that the following levels of attributes are considered relevant and interesting by the marketer for the study :

	Attribute	Levels
1.	Brand	Kitchen Master Elegant Torrent Evermaid
2.	Price	Rs. 15,000/- Rs. 20,000/- Rs. 25,000/-
3.	Washing Capacity	High Medium Low
4.	Colour	Steel grey White Light Blue Pink
5.	Shape	Cylindrical Box Type

Since the 5 attributes can take 4, 3, 3, 4 and 2 levels, the total number of possible product concepts that can be generated by configuring these attributes is 4x3x3x4x2 = 288. In order to determine the part worth utilities of each of the levels of all these attributes, we shall have to take 288 different product concepts for getting his responses. This number is certainly too large for any consumer. Therefore, we resort to the method of Fractional Factorial Design of Experiment to make it manageable.

The statistical technique of Fractional Factorial Design of Experiment finds out the minimum number of product designs which are necessary to use in the study and yet provide us all the information that we originally sought. These designs are also mutually independent (orthogonal) to avoid any redundancy in the data and allow the representation of each of the attributes and their respective levels in an unbiased manner.

In the example of dish washer considered here, this technique has given us only 16 designs out of the 288 possible dish washers. However, it should be noted that such reduction in number of product designs is possible only after making certain assumptions. For example, we had assumed that none of the attributes interact among themselves. Or in other words, the attributes are considered to be independent of each other. Only under this assumption we got the number of product concepts as 16. At the other end, if we would have allowed all the attributes to interact with each other the required number of product concepts would have remained as 288. With different types of assumptions the number of concepts required would be in between these extremes.

The 16 product concepts found through this method are not unique. Many other sets of 16 cards would have also been equally good. However, all of these sets would have to be independent and represent all the attributes and their respective levels in an unbiased manner. We are illustrating below one such set of 16 cards representing the product concepts of dish washers using Fractional Factorial Design of Experiment.

Table 1 : Description of Product Concept Cards.

Card #	Brand	Price	Cap.	Colour	Shape
Card 1	Kit Master	Rs.15,000	High	St. Grey	Cylinder
Card 2	Elegant	Rs.25,000	Medium	St. Grey	Box Type
Card 3	Torrent	Rs.20,000	Medium	St. Grey	Box Type
Card 4	Ever Maid	Rs.20,000	Low	St. Grey	Cylinder
Card 5	Kit Master	Rs.20,000	Medium	White	Box Type
Card 6	Elegant	Rs.20,000	Low	White	Cylinder
Card 7	Torrent	Rs.25,000	High	White	Cylinder
Card 8	Ever Maid	Rs.15,000	Medium	White	Box Type
Card 9	Kit Master	Rs.25,000	Low	Light Bl	Box Type
Card 10	Elegant	Rs.15,000	Medium	Light Bl	Cylinder
Card 11	Torrent	Rs.20,000	Medium	Light Bl	Cylinder
Card 12	Ever Maid	Rs.20,000	High	Light Bl	Box Type
Card 13	Kit Master	Rs.20,000	Medium	Pink	Cylinder
Card 14	Elegant	Rs.20,000	High	Pink	Box Type
Card 15	Torrent	Rs.15,000	Low	Pink	Box Type
Card 16	Ever Maid	Rs.25,000	Medium	Pink	Cylinder

4. Physical design of stimuli

After selecting the product concepts required for the Conjoint Analysis study, they need to be exposed to the consumers as stimuli. This may be done in a variety of ways mainly depending on the demands of the situation and the convenience of the researcher. Of course, it would be most desirable to present real life prototypes of the products according to the product concepts specified. These, may be given to the consumers for their usage or trials. But, such extreme ways of presenting the products may not always be possible or even necessary. In such cases, product models, diagrams or even verbal descriptions may be adopted. In our example of dish washers, it may not be possible to produce the 16 prototypes and take them to the consumers. Just their models or pictures may be sufficient.

5. Data Collection

Ease of data collection is a key feature of Conjoint Analysis. The consumers are asked only to assign rating scores to each of the product stimuli or even rank the different concepts presented to them. This is quite a realistic task and is close to the shopping experiences where the customer merely makes choices. He does not have to respond to each of the attributes separately.

This feature of conjoint analysis is possible due to the use of Fractional Factorial Design of Experiment before collection of data and the use of Conjoint Analysis after collecting the data. In other words, the use of the technique eases the burden of the respondents.

6. Determination of part worth utilities

The rating or ranking data obtained from the consumers are analysed next. Two methods are more popular for this purpose. In one method, the part worth utilities for each of the levels of each attributes are arbitrarily assigned. Based on these assumed values, consumers overall rating or ranking (as the case may be) are estimated. These estimated responses may, understandably, be quite different from the actual data. After a few iterations convergence is achieved so that the part worth utilities found approximate the estimate responses to the actual data best.

In the alternative method, the part worth utilities are derived in one step. Here, an error function describing the difference between the estimated and actual data is defined. This function is then minimised.

After using any of the available method, the output is obtained for each of the respondent separately. This is quite significant as the disagregate data can be combined in any of the desired way. But, if the output was only at the aggregate level then disaggregation might not have been possible.

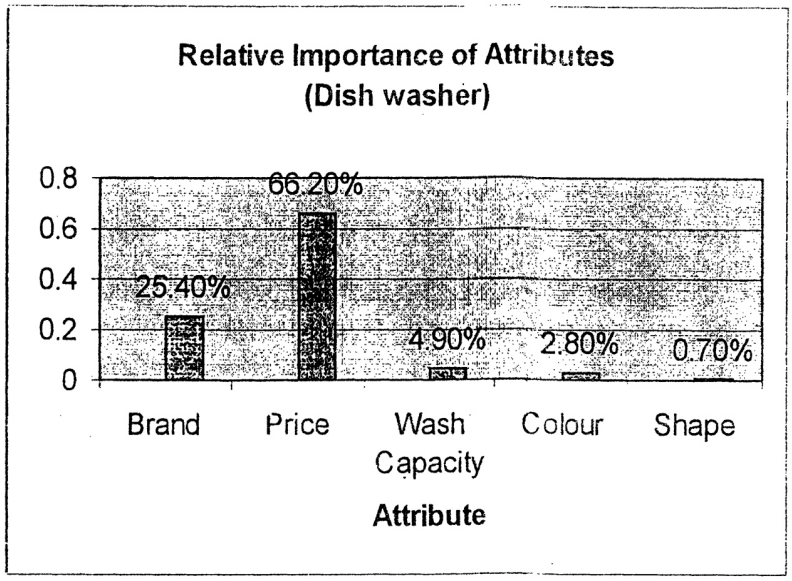
In our example of dish washer, the part worth utilities may be found for each of the attributes and their levels as following:

	Attribute	Levels	Part Worth Utility
1.	Brand (25.4%)	Kitchen Master	2.2
		Elegant	-2.3
		Torrent	-0.8
		Evermaid	0.9
2.	Price (66.2%)	Rs. 15,000/-	5.9
		Rs. 20,000/-	-0.1
		Rs. 25,000/-	-5.8
3.	Washing Capacity (4.9%)	High	0.0
		Medium	0.1
		Low	-0.1
4.	Colour (2.8%)	Steel Grey	-0.3
		White	-0.1
		Light Blue	0.1
		Pink	0.2
5.	Shape (0.7%)	Cylindrical	-0.1
		Box Type	0.1

From the above table, we find that price plays most important role (66.2%) in the minds of customers. This is followed by Brand (25.4%), Washing Capacity (4.9%), Colour (2.8%) and Shape (0.7%). These relative importance values for the attributes have been found by finding the difference between the maximum and minimum values of the part worth utilities of the respective attributes.

The relative importance of different attributes has been shown in figure 1 below.

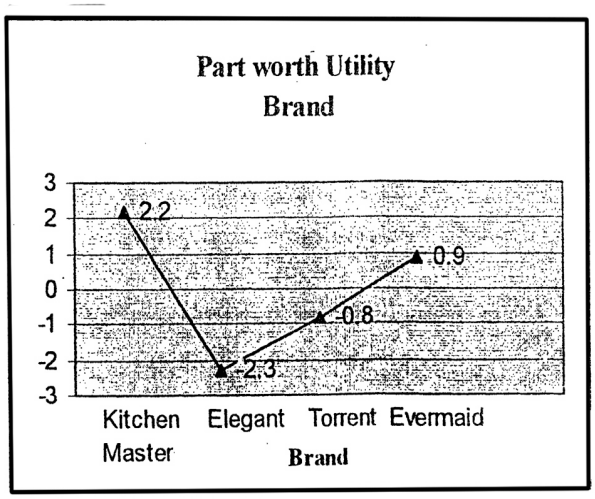
Figure 1 : Relative importance assigned by the customers by the customers to attributes of dish washer

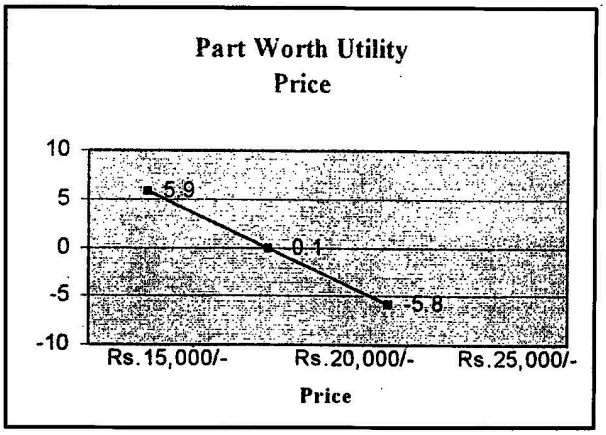


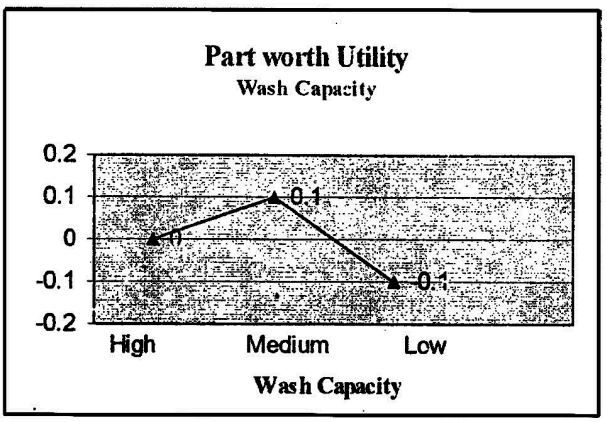
Customers respond quite predictably towards different price levels. They prefer lesser price to the higher ones. Among brands, we find that Kitchen Master is the most liked brand and Elegant is the least liked one. However, they do not prefer the "High" capacity most. They prefer "Medium" size most. In terms of the colour, they prefer Pink colour most. Between the two shapes, box type dish washers are preferred more. But, the importance scores for the wash capacity, colour and shape themselves are quite low.

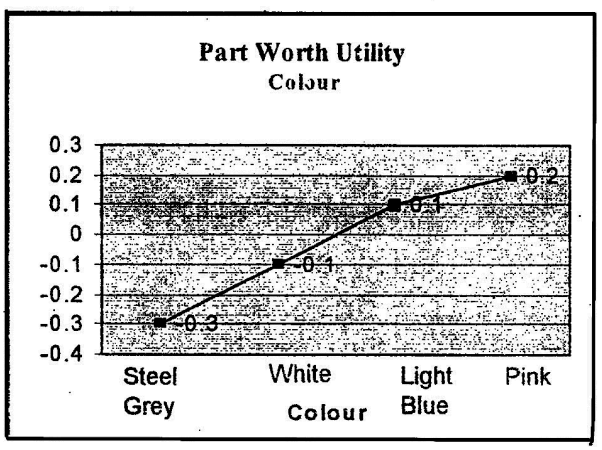
The part worth utilities of conjoint analysis can be represented graphically as shown in the figure 2.

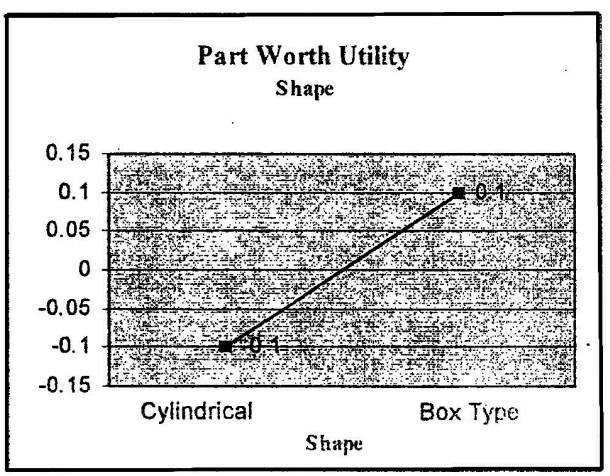
Figure 2 : Part worth utilities assignment by the customers to different attributes levels of dish washer




Part Worth Utility
Price
10
5
0
-5
-10
5.9
0.1
-5.8
Rs.15,000/-
Rs.20,000/-
Rs.25,000/-
Price


Part worth Utility
Wash Capacity
0.2
0.1
0
-0.1
-0.2
0
0.1
-0.1
High
Medium
Low
Wash Capacity


Part Worth Utility
Colour
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.3
-0.1
0.1
0.2
Steel Grey
White
Light Blue
Pink
Colour


Part Worth Utility
Shape
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.1
0.1
Cylindrical
Box Type
Shape

In the above figure, the part worth utilities for different attributes and their respective levels have been shown along the common Y-axis. As a result, the part worth utilities become comparable within as well as across different attributes. This way we can understand the trade-offs that the consumers make.

The part worth utilities can now be added to determine the total utility for each of the possible product concepts. This allows us to scan the consumer preference pattern for all of the 288 product concepts although he has been exposed to only 16 of them. We can now also rank all of these 288 product concepts.

7. Conjoint Analysis applications

Calculation of the part worth utilities becomes just the starting point for many interesting applications of conjoint analysis. The important ones among them are described next:

i. **Optimum Product Design:** Since all possible product concepts can be compared after adding their respective attribute levels part worth utilities, it is possible to determine the demand for different products out of any given set of available products in the marketplace. The demand levels can be converted into profit figures as cost of producing and marketing can also be calculated. These cost calculations are possible as the volume of operations and the features of the products are now known. Thus, the optimum product can be chosen from the profits point of view (or any of the other given management's objective). Customers differential rates of purchase of products are also duly considered at this stage.

Quite often, a manager may like to know the effects of slight change in any of the attribute by his own company or the competitor's. Conjoint Analysis allows this kind of "What if" analysis very easily with the data base of part worth utilities. In fact, different kinds of scenarios can be simulated and the manager can optimise not only the product but other aspects of his marketing strategy. Similarly, whenever there is any change in competitor's actions or in the environment a fresh scenario can be drawn for simulation. Of course, the simulation shall be limited to the attributes considered in the analysis. This feature does also help in increasing the shelf life of the conjoint analysis output.

ii. **Market segmentation:** Since the Conjoint Analysis is done at the individual customer level, the individual customer's identity can be retained throughout the analysis. Thus, consumers can be segmented according to their sensitivities to different product attributes.

It is also possible to identify the customers segments which would be attracted most for the proposed product position. This helps in having a focussed matching between the chosen product position and the target customer segment. It can also help in identifying that part of competitor's market which needs to be poached for snatching market share from them. Similarly, the same type of analysis can be done to identify the most vulnerable section of one's own market segment.

Sometimes, an additional product offer appears to be quite attractive. But, this may be at the cost of cannibalisation. Conjoint Analysis can help in estimating the effects of cannibalisation as well. Thus, it helps in maximising net profits of the organisation.

iii. **SWOT Analysis:** First of all, the part worth utility of the brand itself can tell about the relative brand strength. Similarly by looking at the other features of one's own and competitor's offers Conjoint Analysis enables the marketers to conduct his detailed SWOT analysis.

iv. **Estimating Customer Level Brand Equity:** Conjoint Analysis is a good bridge between the consumer level perceptions and the financial worth of the offers. This can be used for estimating the important parameter of brand equity at the consumers level. There is scope of differentiating the "Loyal", "Acceptors" and "Switchers" for more accurate calculations of brand equity.

13.4 SUMMARY

In this unit we have discussed about Conjoint Analysis, which is a multivariate analysis technique introduced to the marketers in 1970's. Conjoint Analysis is basically a data decompositional techniques which tries to plot the output data on the joint space of the importance of each attribute. The important point to note is that the consumer is not asked to assign scores to different attribute separately. The main steps involved in using conjoint analysis include determination of the salient attributes for the given product from the points of view of the consumer, assigning a set of discrete levels or a range of continuous values to each of the attributes, utilising fraction factorial design of experiment for designing the stimuli for experiment, physically designing the stimuli, data collection, conjoint analysis and determination of part worth utilities. The possible application, of conjoint analysis include product design, market segmentation, swot analysis etc.

13.5 SELF-ASSESSMENT EXERCISES

1. Define conjoint analysis and state its potientail applications.

2. Describe some marketing research problems that you feel may be amenable to conjoint analysis and explain how you would use conjoint analysis in these situation.

3. What are the steps involved in conjoint analysis? Explain with the help of examples.

13.6 FURTHER READINGS

Green, Paul E., Tull, S. and Albaum, G. - "*Research for Marketing Decisions*", PHI, N.Delhi.

Parasuraman, A. - "Marketing Research", *Addision-Wesley Publishing Company*, U.S.A.

UNIT 14 CLUSTER ANALYSIS AND MULTI-DIMENSIONAL SCALING

Objectives

After going through this unit on Cluster Analysis, & Multidimensional Scaling you should be able to understand the following :

- Importance of Cluster Analysis in today's market conditions
- Inputs for cluster analysis
- Different approaches to cluster analysis
- Nature of Multi-Dimentional Scaling (MDS)
- Applications of MDS

Structure

14.1 Introduction

14.2 Inputs for Cluster Analysis

14.3 Different Approaches to Cluster Analysis

14.4 Factors to be considered while using Cluster Analysis Techniques

14.5 Multi-dimensional Scaling

14.6 Applications of Multi-Dimensional Scaling

14.7 Summary

14.8 Self-Assessment Exercises

14.9 Further Readings

14.1 INTRODUCTION

Cluster analysis is a technique that is used in order to segment a market. The objective is to find out a group of customers in the market place that are homogeneous i.e., they share some characteristics so that they can be classified into one group. The cluster/group so found out should be large enough so that the company can develop it profitably, as the ultimate objective of a company is to serve the customer and earn profits. The group of customers that the company hopes to serve should be large enough for a company so that it is an economically viable proposition for the company. This is also true for the customer as customer would not be willing to pay beyond a certain price for a particular product (price of course is a function of positioning of product, cost of production etc.).

As an example, let us consider the Watch Industry. There could be many ways in which the Watch Industry could be segmented which are as follows :

a) Gender (Male/Female)

b) Technology (Digital/Analog)

c) Design Features

d) Occasion of Use (Formal/Casual/Party)

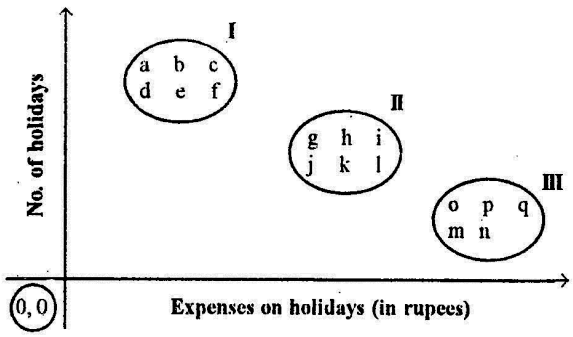
e) Price (Low/Medium/High/Jewellery)

Some of the above segmentation factors are demographic (price, gender) whereas some are psychographic factors (occasion to use.)

This, therefore, presents a problem to the market researcher/company, as to how to identify combination of factors that can be used to segment the market place. It is not always

possible to segment a market on the basis of one single factor. Thus, a combination of factors must be used to segment the market place. And this is where Cluster Analysis technique specifically deals with how objects (people, places, products) should be assigned to groups, so that there should be similarity within the groups, and as much difference between the groups, as possible.

Example



In the above example, cluster analysis has identified three distinct clusters (groups) as follows:

Group I : People who spend less money but have more number of vacation days. (adventure seekers, students)

Group II : People who spend average amount of money and have average number of vacation days. (1st class train travel, 3-star hotels etc.)

Group III : People who have very less vacation days, but spend heavily in these few days (five-star luxury, air travel, foreign trips etc.).

These grouping would help a travel and tourism company in the following ways:

a) The travel and tourism company can study the various clusters of customers that have emerged, and decide which customer group it would like to serve, depending upon the company's own resources and capabilities, the volume of business in each cluster group that will generate sufficient business for the company's own survival.

b) Once a cluster has been selected by the company, it can tailor various tour programmes for its cluster of customers.

c) The company can keep profile of its customers, and identify any new emerging group of cluster.

d) The company can decide to serve either only one group of customers (Niche marketing), or serve all groups of customers, at one and the same time, but having a range of travel and tour programmes.

14.2 INPUTS FOR CLUSTER ANALYSIS

Since cluster analysis is used for the purpose of segmentation of the market place, for using the cluster analysis technique, one has to go through the following procedure on a very broad level :

i) Company identifies its own product capabilities, financial strength, distribution network (Who am I?)

ii) Company identifies the basis on which the customers are to be clustered e.g., one factor segmentation (age, income, gender, geographic location) or two-factor basis of segmentation of the market place (age and income, quality and price, nutrition value and price), or multi-factor basis of segmentation of the market place (Who am I going to serve?).

The company/market research agency has to decide which of the attributes are to be employed that would generate natural grouping in the market place. The choice of variables that are

to be used to cluster the objects is one of the most critical decisions analysis make. The best way is to select those variables that make sense conceptually, rather than using any or all variables. Ideally, the variables should be selected within the framework of an explicitly stated theory that is used for supporting the classification.

iii) Respondents who are part of the research work are given to fill up questionnaire. The questionnaire is typically Interval scaled (5pt, 7pt, 10pt) questionnaire.

The cluster analysis program that is run on the computer identifies the different groups (clusters) or respondent groups such that the members of any one group (cluster) are similar to each other, but different from the respondents in the other groups. The computer program for cluster analysis searches through the answers that the respondents give to the various questions in the questionnaire, and groups together those respondents who have given identical or almost similar answers of a certain combination of the questions. These respondents whose answers are similar are formed into one cluster. The procedure is then continued till more cluster are formed.

14.3 DIFFERENT APPROACHES TO CLUSTER ANALYSIS

There are various approaches adopted for cluster analysis

One extreme approach is to cluster respondents in terms of all available information about the respondent - although this is very thorough, but may lead to difficulty in interpretation and marked difference between groups i.e., the clusters or groups of respondents that are produced may be a hotch-potch, rather than clearly marked description of the target population.

Another extreme approach involves grouping respondents only on data that is directly representing the specific consumer need. Example - quality, price of a product. This has the advantage of clarity and accuracy, but may be quite limited in describing the target population.

The way one approaches cluster analysis depends upon the objective of study. Suppose the objective of the company is physical development of an existing product, then the consumers needs regarding the product are the most suitable variables, for forming clusters.

Suppose an advertisement for a product is to be developed, then the attitude of the consumer towards the product gains more importance, as there are groups of people who have positive, negative or neutral attitude towards product/company. And the advertisement developed would be targeting specific clusters of consumers, depending upon the product/communication need.

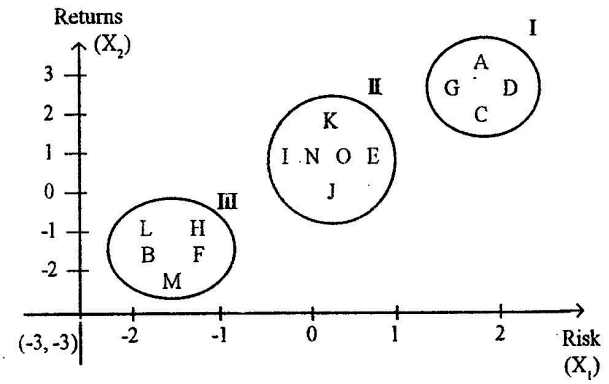
Suppose a company wishes to develop a new product concept/product changes in the existing product offering, then also attitude of consumer towards product/company/product category gains more importance over other attributes when developing clusters.

Therefore, depending upon the objective of study, various characteristics can be used to describe the consumer, his attitudes, needs, demographic characteristics.

A general principle followed in cluster analysis is to omit behavioural and brand image data. One should concentrate on characteristics describing the consumer and the circumstances e.g., specific needs, demographic variables, general attitude and select any variables which would interact with each other to influence consumer behaviour. Once attributes are identified on the basis of which one is trying to form natural groupings of consumers, one has to scale the attributes, so that one has a basis to measure the similarity/differences between/within groups.

Example :

Suppose one wants to understand the consumers who operate in the finance market. With regard to the finance market, there are few outstanding aspects like Risk, Returns, Liquidity. One could use any two aspects at one time to form clusters of consumers, or could use more than two attributes also.



There are 15 consumers, whose responses have been analyzed with respect to Risk and Returns in the financial market.

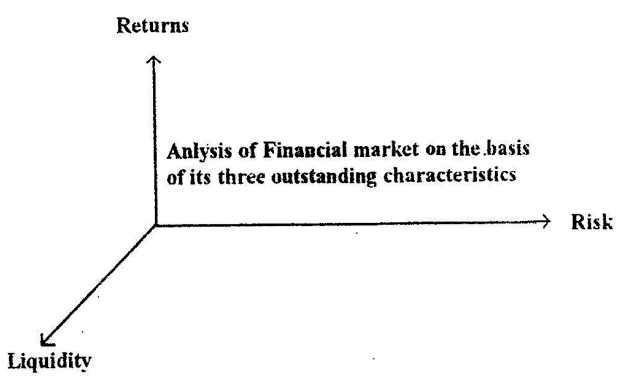
- Cluster I is a group of consumers who believe in High Risk High Returns instruments e.g., Equity
- Cluster II is a group of consumers who believe in Average Risk Average Returns (Mutual funds, Company fixed deposits)
- Cluster III is a group of consumers who beiieve in Low Risk Low Returns (savings bank, fixed deposits).

An obvious way to measure similarity is to calculate the Euclidean distance between points. In the above figure, the distance between points A, and C would calculated as

$$^{d}(A,C) = \sqrt{(Xc_1 - XA_1)^2 + (Xc_2 - XA_2)^2}$$

And so on, one could calculate distance between various points. In three dimensions, the expression for distance between A and C would be

$$^{d}(A,C) = \sqrt{(Xc_1 - XA_1)^2 + (Xc_2 - Xc_2)^2 + (Xc_3 - XA_3)^2}$$



Distance Risk

Returns	1	2	3	-	-	-	-	15
	A	B	C	-	-	-	-	O
1	A							
2	B							
3	C							
-	-							
-	-							
15	O							

The distance between any two points is an inverse measure of similarity. Thus, larger the distance between two points, greater is the difference or dissimilarity between the two points. The closer the distance between two points, the more the similarity between them. Also more the number of respondents to be classified into clusters, more will be the number of distance calculations. Hence for large sample sizes of respondents, computers would have to be used.

After calculating distances, one has to decide how to form clusters. There are various methods of doing it as follows :

i) Nodal Method

In this method one selects a point or points that will serve as nodes for forming clusters or focal point. The rest of the points are then allocated to each cluster on the basis of their distance from the focal point.

The above method is demonstrated as follows:

a) Choose as nodes those objects (points) that have the greatest distance between them.
b) Consider these two objects as focal nodes and allocate all the remaining objects to one cluster or the other based on their distance from the polar nodes.
c) Split the resulting two clusters in the same way. Continue this process till the collection of points is split into its original members.

In the above example, the points 'D' and 'M' have the biggest distance between them, and they will be considered as nodes for the two clusters. Each of the remaining objects would then be allocated to each cluster on the basis of the shortest distance to either D or M. Thus the original cluster would be :

Group 1 : DACE GK
Group 2 : MBFH IJLNO

In Group 1, the least similar (greatest distance) are D and K. They would thus be considered as new nodes, and the items in Group 1 would be allocated to each of the new clusters on the basis of their distance from these new nodes. The new groups would be :

Group 1A : DACG
Group 1B : KE

Similarly, Group 2 would be divided (using B and J as new nodes) to yield

Group 2A : BFHLM
Group 2B : JINO

Thus, after every such stage, one has to find out if the sub-groups that have emerged can be further combined on the basis of some measure of similarity between the objects.

ii) Factor Analysis Method

Another way of developing clusters is by using the method called as Q-factor analysis. By this method one can determine which objects logically belong together. This is also called as inverse factor analysis.

iii) Linkage Method

There are three methods in this - Single Linkage Method, Complete Linkage Method and Average Linkage method.

Using the previous example of analyzing financial markets, the method starts by finding out the points with shortest Euclidean distance. Therefore, initially the following groups emerge:

C-D - Group 1
A-G - Group 2
N-O - Group 3
FM - Group 4

In the next stage, depending upon the cut-off distance, remaining objects are associated to the above clusters. The final clusters that emerge are :

Group 1 - AGDC
Group 2 - NOKEIJ
Group 3 - MFLBH

The complete linkage option starts out in the same manner by clustering the two closest points. However, the distance between two clusters is the longest distance from a point in the first cluster to a point in the second cluster.

In the average linkage option, the distance between the two clusters is the average distance from points in the first cluster to points in the second cluster.

Activity 1

Study the products and the consumers of your company (or the company you are familiar with) and try to develop cluster of customers depending on the different requirements of the costomers.

.....

.....

.....

.....

.....

14.4 FACTORS TO BE CONSIDERED WHILE USING CLUSTER ANALYSIS TECHNIQUE

The following aspects should be kept in mind while using cluster analysis method :

a) A number of clusters may emerge after doing a cluster analysis. However, there is a limit to the number of clusters that a company can consider due to
   - limitation of market potential within a cluster
   - differences between clusters not sharply defined

b) Cluster analysis provides a way of segmenting the market but these segments are

not water tight compartments. Products that are developed for a particular segment may attract people from other segments too.

c) The characteristics of a cluster may change over time, as the consumers economic status, education, lifestyle, etc., change over time then the company has to take a relook at the market place.

d) The clusters that have been identified are used developing further marketing strategies in the areas like product developments, advertising research, distribution strategies, pricing strategies etc.

e) The most important assumption in cluster analysis is that the basic measure of similarity on which clustering is based is a valid measure of the similarity between objects. A second major assumption is that there is theoretical justification and basis for structuring objects into clusters. As with other multivariate techniques, there should be theory and logic underlying the cluster analysis.

f) The major limitation of cluster analysis is the difficulty in evaluating the quality of the clusters. It is very difficult to know exactly which clusters are very similar and which objects are dissimilar, and also difficult to select clustering criterian.

In conclusion, cluster analysis is a scientific method of help understand the consumer groups with their differing needs and perceptions.

14.5 MULTI-DIMENSIONAL SCALING

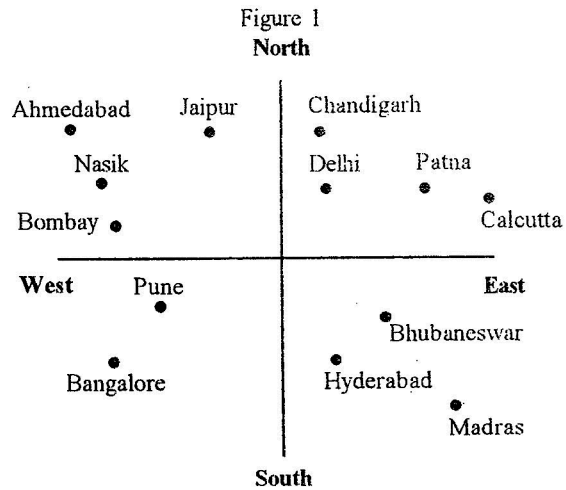
Multi-dimentional Scaling (MDS) technique or perceptual map or positioning map is a multi-variate analysis technique that is particularly suitable for measuring human perceptions and preferences for products/brands. MDS is concerned with spatial representation of relationships among behavioural data i.e., buyer perceptions and preferences.

MDS Fundamentals

a) Metric MDS
b) Non-Metric MDS

Consider the following geographic location of various cities in our country:

Figure 1



The distance between any two points on the map is given by the formula

$$d_{12} = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

The distance between various cities could be calculated and put in a tabular form as follows:

Figure 2 Distances in Kms.

Cities	Bombay	Pune	Nasik	Ahmedabad	Delhi
Bombay					
Pune					
Nasik					
Ahmedabad					
Delhi					

In short, it is a relatively simple matter to go from the map in figure 1 to the set of numerical distances in figure 2.

However, the converse or going from figure 2 to figure 1 is very difficult. Given different objects and their attributes, every consumer perceives the objects in a different manner.

So, one has to first understand the perceptions (distances) in the mind of the consumer regarding products, brands, and plot a graphical representation for the same. The task is to work backwards. That is, we would have to find out simultaneously, the

- Number of dimensions (can be more than two-dimension perceptual mapping)
- Configuration (or pattern) of points in that perceptual space, so that their computed inter-point distances most closely match the input data of figure 2. This is the problem of metric MDS. Whenever input data is interval-scaled or ratio-scaled, it is called as Metric MDS.

The same input data in the format of figure 2 could be transformed into a rank order. Suppose we take the smallest distance between two cities and give in the Rank 1. Then we could arrange the remaining distances between cities in ascending order and give it ranks from 2 onwards up to Rank 13 (e.g., largest distance could be between cities Calcutta and Bombay on the map. We therefore find with this method.

- Number of dimensions
- Configuragion (pattern) of points in that dimensionality, so that the ranks on their computed inter-point distances most closely matched the ranks of the input data. This is the problem of non-metric MDS. Whenever input data is in the form of norminal or ordinal form, it is called as non-metric MDS.

Whichever method for MDS one select (metric and non-metric MDS), it does not change the relative distances of the points.

Attribute Space — Objective Space and Perceived Space.

Any product or service can be visualized as composed of both objective and perceived attributes or 'dimensions'. A company has an objective description of its product in physical or chemical terms. Such an objective attribute 'space' in which various brands are viewed as points positioned in space, will usually not agree with the buyers perception of the products. The buyers perception about a product, service, company is a sum total of not only the company which is selling the product, but also other factors in the market like competition, competitor's products advertising, after-sales service etc.

Also note that the dimensions of perceived space need not be the same as the objective space. As an example, for many consumer goods that are highly individualistic items like tea, coffee, toothpastes, perfumes, shampoos, where strong individual preferences play a key role in buying decisions, the objective space of the company and perceived space of the consumer may be widely different. But in the case of industrial products, the perceived configuration of brands may agree rather closely with an objectively constructed configuration in which

measurements of such characteristics as speed, reliability, efficiency can be made rather straightforwardly.

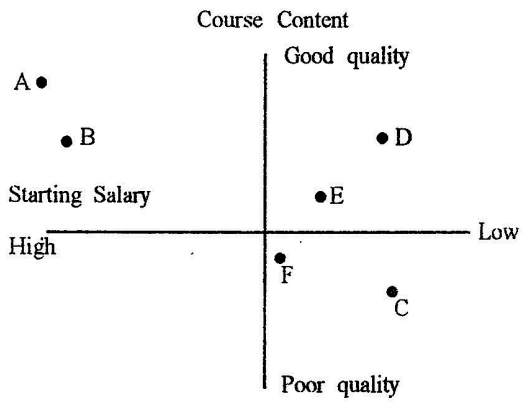
Ultimately, the view of the consumer is the most relevant one, as this is what would make the consumer go out and buy a particular product/brand.

Example of Multi-Dimensional Scaling

A study was conducted in a leading business school (A) among the first year MBA students to find out how other business schools were 'similar' or 'different from the 'A' business school. Various attributes are used to evaluate a business school - faculty, faculty/student ratio, starting salaries of the MBA students, physical facilities like library, computers, and so on. But which of the attributes do students 'actually' use in making a competitive comparison between the business schools?

Let us consider six leading business schools in a place as A, B, C, D, E, F. Around hundred students were given a questionnaire in which the students were asked to respond to the questions which indicated various attributes of a management school, and students had to mark schools that were similar/dissimilar on the different counts.

From the information that was gathered from the questionnaire and subjected to MDS software program, what emerged was that course content and starting salary are the most important criteria by which students judge a business school.



The multi-dimensional scaling technique is therefore very useful in plotting the perceived perception of the consumer about different brands/products/services. And this analysis could be used by the marketing manager of a company for developing further marketing strategies.

Activity 2

Consider the following products and try to develop a perceptual map of the various brands in that particular product category. Before attemting such an exercise, the student should ,understand the product category in its totality i.e., target consumer, pricing strategy advertisement strategy, distribution, usage and attitude for the product, and then try to develop a positioning map (MDS) for the same.Once a product has been studied, the study should try to identify two dimensions along which brands could be clubbed as similary dissimilar

1) .Toothpaste (Colgate Gel, Colgate Total, Pepsodent 2-in-1, Babool, Close-up etc.)
2) Shampoo (Nyle Herbal, Dabur Vatika, Sunsilk, Lakme, Clinic ets.)

.....

.....

.....

.....

.....

14.6 APPLICATIONS OF MULTI-DIMENSIONAL SCALING

Some of the typical marketing applications that emerge from the MDS technique are :

1) Market Segmentation

Market segmentation is the technique of trying to identify groups of consumers who exhibit commonality of perception of products and preferences. One can use MDS techniques to identify present perceptions of products by consumers, and use it modify the company's product, package, advertising, additional features, so that the product offering of the company moves more and more closer to the 'ideal' requirement of the consumer.

2) Advertisement Evaluation

The MDS technique could be used at the stage of advertisement pre-testing. Once an advertisement has been developed, it could be tested for similarity/dissimilarity with other advertisements in the same product category. As the ultimate objective of an advertisement is to communicate with the target consumer effectively, and this is possible only if the advertisement is distinct in its message from the other competing advertisements.

3) Product Re-positioning Studies

If a company is interested in re-positioning its product/service (in the mind of the consumer), the first and foremost activity to be done is to assess the current perception of the product in the mind of the consumer. The classic re-positioning case is that of Cadbury chocolates, which kept on assessing its positioning platform, and successfully moved Chocolates from a product perceived as one for children, to a product which could be consumed by a person of any age, at any time, of the day, and for varied occasions.

4) New Product Development

MDS technique shows us the various perceived perceptions of the different brands. Spaces/ Gaps in the product perceptions could be used to develop new offerings for the target consumer.

5) Test Marketing

MDS technique can be used to identify cities that have similar demographic characteristics, and one could then identify a city which could represent a national character, and use that city for test marketing.

One can thus observe that MDS is a very useful technique to help understand the market place and develop strategies for the future.

Advantage of MDS

The advantage of MDS methods is not in the measurement of physical distances, but rather "psychological distances", also called as 'dissimilarities'. In MDS, we assume that every individual person has a 'mental map' of products, people, places, events, companies; and individuals keep on evaluating their external environment on a continuous basis.

We also assume that the respondent is able to provide either numerical measure of his or her perceived degree of similarity/dissimilarity between pairs of objects, or can rank pairs of objects (ordinal scale of measurement) in terms of similarity/dissimilarity to each other.

We can then make use of methodology, of MDS to construct a physical map in one or more dimensional whose inter-point distances (or ranks of distances) are most consistent with input data.

Now-a-days a number of software programmes are available for conducting MDS analysis. These programmes provide for a variety of input data. Some of the widely used softwares include MDPREF, MDSCAL SM, INDSCAL, PREFMAM, PROFIT, KUST.

14.7 SUMMARY

In this Unit we have discussed two of the techniques of multi-variate analysis viz., 'Cluster Analysis' and Multi-Dimensional Scaling. Cluster Analysis is mainly used for market segmentation. Some of its other applications include test market selection and to establish grouping of products within a product line. Various inputs required for cluster analysis have been discussed in the Unit. Further various application to cluster analysis have been explained. Finally some of the important consideration which should be kept in mind while using cluster analysis have been highlighted. In this Unit we also discussed the Multi-dimensional Scaling technique or perceptual map or positioning map is a multi-variate technique that is particularly suitable for measuring human perceptions and preferences for products/brands. Its applications include market segmentation, positioning studies. test marketing etc.

14.8 SELF-ASSESSMENT EXERCISES

1) What is cluster analysis? What are its possible applications?
2) State how the following techniques differ from each other (a) Cluster Analysis (b) Discriminant Analysis.
3) An airlines marketing manager wants to segment his customer. How can cluster analysis be used for this purpose?
4) Give few examples of marketing situations where cluster analysis can be used.
5) Discuss with the help of examples the areas where Multi-dimentional Scaling can be applied for marketing.

14.9 FURTHER READINGS

Green, Tull and Albaum "*Research for Markeint Decision*", PHI. N. Delhi.

Majumdar, R. "*Marketing Research: Teat, Applications and Case Studies*", Wiley, New Delhi.

Beri, G.C. "*Marketing Research*", Tata McGraw Hill Publishing Company Limited, New Delhi.

Parasurama, A. "*Marketing Research*", Addison - Wesley Publishing Company Inc., U.S.A.

UNIT 15 APPLICATION OF MARKETING RESEARCH IN INDIA - SOME CASE STUDIS

Objectives

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

- appreciate some of the selected areas in which the techniques of marketing research are being applied

Structure

15.1 Introduction

15.2 Illustration on 'Consumer Research'

15.3 Illustration on 'Demand Forecasting'

15.4 Illustration on 'Sales Promotion Campaign'

15.5 Other Potential Applications

15.6 Summary

15.7 Self-Assessment Exercises

15.8 Further Readings

15.1 INTRODUCTION

In the last few units we have given an indeptch background of the univariate and multivariate procedures for analysing marketing research problems. Let us now address ourselves of the task of this unit, by seeing how far the techniques of marketing research are effectively put to practice in India. what are their common applications?; what are their potential applications?

Our selected applications would include (1) 'Consumer Research' — a study on refined oils, (2) 'Demand Forecasting Forecasting Demand for light Commercial Vehicles (LCV), and (3) 'Sales Promotion Campaign' — a campaign successfully launched by 'Richardson Hindustan' (now Proctor & Gamble India Ltd.)

We would like to highlight the other potential applications of marketing research methods in the areas of new product introduction, advertising and media planning.

15.2 ILLUSTRATION ON 'CONSUMER RESEARCH'

Consumer Research is one area where the marketing research firms are very active. A number of market surveys are being carried out at regular intervals both by the agencies as well as by big organisations selling consumer products. The focus of any consumer survey is to find out 'what does the consumer want?' 'What are his preference', 'his own perception of the product', 'his loyalty' etc. Consumer Research is a powerful aid in the introduction of any new product. It also provides a database to understand the consumer buying behaviour and enables organisations to take remedial actions in regard to marketing strategies when called for.

The present illustration is concerned with a market study on 'refined oil' in New Delhi carried out a couple of years ago as an exploratory research study.

Research Objectives

1 To assess the awareness level of the existing brands of refined oil by the consumers and also to identify the source of awareness.

2 To analyse the consumption pattern of cooking medium with particular emphasis on refined oil.

3 To analyse the purchase pattern of refined oil in terms of 1 Kg. tin, 2 Kg. tin, etc. and also loose purchasers.

4 To analyse and identify important factors that are considered while purchasing any refined oil.

5 To analyse the consumer perception of the important attributes in respect of the various brands of refined oil.

6 To test for the statistical hypothesis of no association between brand preference and income level.

Methodology

A probability sampling had been adopted covering about 160 households (161 to be exact) in all important areas of Delhi. Data collection involved a questionnaire (specimen copy attached) to meet the research objectives.

The questionnaire was administered through personal interview in the households allowing full participation from the respondents.

Data analysis, interpretation of results and conclusions, drawn from the study are being discussed here.

Data Analysis

1 Awareness of brands as revealed from memory recall.

No. of respondents = 161

Name of Brand	Frequency of Awareness							
	Income Groups							
	(Upper income) (44)		Middle income (77)		Lower Middle income (40)		Total (161)	
	No. of Respondents	%	No. of Respondents	%	No. of Respondents	%	No. of Respondents	%
Postman	38	86.4	42	54.5	34	85.0	114	70.8
Dalda	19	43.2	53	68.8	19	47.5	91	56.5
Ruby	28	63.6	32	41.6	17	42.5	77	47.8
Diamond	23	52.3	33	42.9	14	35.0	70	53.5
Suffola	25	56.8	24	31.2	11	27.5	60	37.3
Cornola	8	18.2	2	2.6	3	7.5	13	8.1
Cornello	3	6.8	–	–	–	–	3	1.9
Markfed	2	4.5	–	–	1	2.5	3	1.9
Others	1	2.3	–	–	7	17.5	8	5.0
Total reactions	147		186		106		439	

Figures (44), (77) and (40) indicate number of households or respondents interviewed. The percentage awareness is worked out to these figures.

Note : Because one respondent can recall more than one brand and even than 3 or 4, the percentage are not additive.

Distribution of awareness to Total Frequency.

Name of brand	No. of times a brand is mentioned	% of total no. of calls
Postman	114	25.97
Dalda	91	20.73
Ruby	77	17.54
Diamond	70	15.95
Suffola	60	13.67
Coronola	13	2.96
Cornello	3	0.68
Markfed	3	0.68
Others	8	1.82
Total No. of calling	439	100.00

2 Source of awareness : Total Respondents = 161

Source	No. of times each source is mentioned	% of total no. of frequency	% to Total Response
Advertisement	80	40.61	49.69
Friends	55	27.92	34.16
Market/Shop	18	9.14	11.18
Personal use	33	16.75	20.50
Others	11	5.58	6.83
Total	197	100.00	

3 Consumption Pattern

Monthly Consumption in Kg.

Income Class	Refined Oil	Vanaspati/Ghee	Raw Oil
Lower middle income	82.5	104.00	35.5
Total family size = 203			
Per capita consumption	0.41	0.51	0.17
Middle income	164.00	177.5	35.3
Total family size = 368			
Per capita consumption	0.45	0.48	0.10
(Upper Income)	145.50	95.5	29.0
Total family size = 149			
Per capita consumption	0.98	0.64	0.19
Overall	392.0	377.0	99.8
Total family size = 720			
Per capita consumption	0.54	0.52	0.14

4 Purchase pattern (No. of households buying)

Income group	Loose Buyers	1 Kg	2 Kg	4 Kg	Bulk (15.5 Kg or 16 Kg)	Total %
Lower middle income	6 (24)	3 (12)	8 (32)	5 (20)	3 (12)	25 (100)
Middle income	16 (23)	6 (9)	25 (36)	20 (28)	3 (4)	70 (100)

Upper income	8 (23)	4 (11)	6 (17)	13 (37)	4 (12)	35 (100)
Total	30 (23)	13 (10)	39 (30)	38 (29)	10 (8)	130 (100)

5 Rank distribution (based on 130 households consuming refined oil out of 161)

Particulars	Weighed Average Rank
Odourless	2.37
Taste of food cooked	2.52
No cholesterol	3.21
Purity	3.29
Good for heart	4.22
Price	5.30
Colour	6.55

6 Rating of attributes for different brands on a '0' to '10' Scale.

	Postman	Dalda	Ruby	Doamond
Odourless	8.26	8.42	7.60	7.53
Taste of food-cooked	8.44	8.02	8.75	7.59
Purity	8.45	7.88	8.55	7.19
Good for heart	7.69	7.50	6.46	5.43
No smoke	8.24	7.48	7.40	6.44
No cholesterol	7.70	7.51	6.65	5.44
Total observations on which weighed average is worked out	58	43	20	32

7 Testing of Hypothesis set up in our problem dimension

H_0 : Brand preference and income levels are independent (i.e. not associated)

H_1 : Brand preference and income level are associated.

Cross classification data (Figs. in each cell indicated no. of persons using a brand)

	Income level/Month			
Brand	upto 2000	2000-3000	Above 3000	Total
Postman	18	17	13	48
Dalda	3	11	19	33
Ruby	2	8	4	14
Diamond	6	22	4	32
Total	29	58	40	127

$\chi^2 = \Sigma (O_{ij} - E_{ij})^2/E_{ij}$ where O_{ij} is the observed frequency in the i, jth cell and E_{ij} is the expected frequency in the i, jth cell.

We calculate

$$\chi^2 = \frac{(18-11)^2}{11} + \frac{(3-8)^2}{8} + \frac{(2-3)^2}{7} + \frac{(6-7)^2}{7} + \frac{(17-22)^2}{22} +$$

$$+ \frac{(11-15)^2}{15} + \frac{(8-6)^2}{6} + \frac{(22-15)^2}{15} + \frac{(13-15)^2}{15} + \frac{(19-10)^2}{10} +$$

$$\frac{(4-5)^2}{5} + \frac{(4-10)^2}{10}$$

= 26.17 degrees of freedom = $(4 - 1) \times (3 - 1) = 6$

χ^2 at 5% level from table = 12.592
Reject H_0 and accept H_1

Let us also calculate χ^2 considering only the income levels 2000-3000 and above 3000. In this case
degrees of freedom = $(4-1) \times (2-1) = 3$

$$\chi^2 = \frac{(17-18)^2}{18} + \frac{(11-18)^2}{18} + \frac{(8-7)^2}{7} + \frac{(22-15)^2}{15} + \frac{(13-12)^2}{12} + \frac{(19-12)^2}{12} + \frac{(4-5)^2}{5} + \frac{(4-11)^2}{11} = 15.01$$

χ^2 at 5% level from table = 7.815
Reject H_0 and accept H_1

Interpretation of results (based on data analysis)

Awareness

1 The study revealed that Postman had the highest level of awareness of 70.8% followed by Dalda of 56.5%, Ruby of 47.8%. Diamond of 43.5% and Suffola of 37.3% CORNOLA, came next in the awareness level of a low 8.1%. This indicated that the probable cause for its poor sales could be lack of awareness. A further perusal indicated that out of the 13 respondents who answered they were familiar with CORNOLA, 8 respondents came from posh locality.

The same trend was seen in the next table where the level of awareness was worked out as % to total frequency (No. of Calls) Postman 25.97%, Dalda 20.73%, Ruby 17.54%, Diamond 15.95%, Suffola 13.67%, and CORNOLA 2.96%.

Therefore one thing was conjectured that the awareness level of CORNOLA which was low could be the probable cause for its poor sales. If better sales performance and additional sales were to be generated, the present level of awareness should be increased to a reasonably higher level.

2 How to increase awareness? This was answered from table 2 which gives sources of awareness, frequency of awareness, % of awareness' source measured to total frequency and to total respondents. Advertisement was maximum 40.61%, followed by friends 27.92%, market shop 9.14% and personal use 16.78%. It was, therefore inferred that effective advertisement campaign could be the best method for increasing awareness. It was felt that it would be a good idea that the advertisement was not only in T.V. but also given in a magazine like 'Femina' which had a larger volume of circulation and was widely read by housewives. The advertisement should highlight the special attributes of CORNOLA.

Before we talk of consumption pattern, it may be mentioned in the passing that only 12 out of 161 respondents answered they were using refined oil on the advice of medical doctor. This was not a significant and encouraging figure. Therefore 'CORNOLA' could not gain much sales on this ground. With regard to the base of the oil namely Groundnut, Corn, or/any other, all people answered it (excepting one or two) correctly.

3 Consumption Pattern

It was significant to note from the table that the monthly per capita consumption of refined oil was increasing from lower-middle income to middle income and upper income group. Let us look at the following table giving the per capita consumption for each group and for each type of cooking medium.

Per capita consumption (Kg.)/Month

	Refined Oil	Vanaspati Ghee	Raw Oil
Lower middle income	0.41	0.51	0.17
Middle income	0.45	0.48	0.10
Upper income	0.98	0.64	0.19

As per capita consumption of refined oil was maximum in upper income group, it was decided to concentrate on a priority basis on this income group, for promoting

UGSEC-02/255

new refined oil like CORNOLA. It may be mentioned in the passing that people are using more and more of refined oil and less and less of ghee in the posh localities.

4 Purchase Pattern

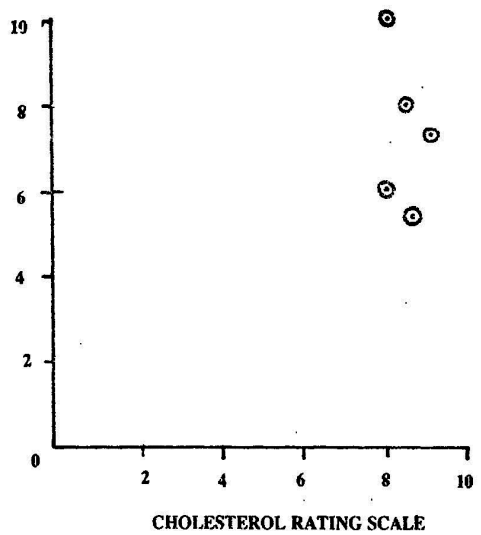
2 Kg. tin and 4 Kg. tin appeared to be more popular in terms of purchase. The purchases were 30% and 29% respectively for 2 Kg. tin and 4 Kg. tin. About 23% were purchasing in loose. It was also revealing to note in posh locality 4 Kg. tin was more popular. (37% purchase). So concentration should be on 4 Kg. tin and 2 Kg. tin with respect to purchase because they were more widely purchased.

5 Important Factors (Ranked)

Odour, taste and cholesterol were the three most important factors that go with the purchase of any refined oil. Certainly if CORNOLA could match with others in respect of odour and taste of food cooked, it would definitely have a differential advantage in respect of cholesterol over the existing brands. This factor should be highlighted more in the advertisement as a special feature. Opinions based on people who used CORNOLA suggested, that with respect to odour and taste, it was equally good if not better than others. Average rank for odour was 2.37, taste 2.52 and cholesterol 3.21 based on 161 observations.

6 Attributes Rating

The rating on 0 – 10 scale revealed again that Postman and Dalda were perceived better than others. Let us draw the perceptual map on the basis of the weighted average scores already worked out for the two attributes 'odour' and 'cholesterol' Let us superimpose CORNOLA in the perceptual map based on opinion expressed by the users who are of course not falling in our sample list. It should ideally score 10 points for cholesterol because it has 0% cholesterol.



*CORNOLA is superimposed in the map. Thought should be given as to how to readjust brand positioning and enable consumer to have proper perception of this product.

7 Hypothesis Testing

i) Taking all income level calculated $\chi^2 = 26.17$ which exceeds table χ^2 at 5% level = 12.592. Therefore reject H_0 which means brand preference is associated with income level.

ii) Taking Rs. 2000 – Rs. 3000 and above Rs. 3000 income calculated $\chi^2 = 15.01$ Table χ^2 at 3 degree of freedom at 5% level = 7.815. Reject H_0 which means again that income level and brand preference are associated.

In both cases we found income level and brand preference were associated and were not independent. This also meant that there was a strong brand loyalty. This would be difficult to break unless followed by aggressive advertising campaign. It was

inferred that there was scope for creating good sales in view of refined oil getting more and more popular, provided effective advertising campaigns were launched which would improve consumer acceptability and awareness.

Concluding remarks

1 Sales of CORNOLA was poor due to lack of awareness of this brand by consumers. Awareness level should be increased by proper advertising campaign including advertising in a magazine like 'Femina'

2 Consumption pattern on the basis of per capita consumption suggested that on a priority basis posh localities should be fully concentrated upon and exploited in respect of promoting 'CORNOLA'.

3 Purchase pattern suggested that the most popular packed tins were 2 Kg. and 4 Kg. CORNOLA should be packed and marketed initially in these sizes.

4 Three important factors for the purchase of refined oil were odour, taste of food cooked and cholesterol. With respect to cholesterol CORNOLA had a distinct advantage because it was cholesterol free. This should be highlighted in all advertisements. Perceptual map of attributes (odour, cholesterol) revealed that CORNOLA could be ideally on top with respect to cholesterol and reasonably good with respect to taste/odour. It should be pushed towards Postman and Dalda by vigorous advertisements.

5 Brand preference and income levels were associated as validated by statistical test of χ^2. This should be countered by vigorous advertisements.

To sum up, a bad product cannot sell however effective the advertisement is, but at the same time a good product cannot sell unless backed up by effective advertisements.

Activity 1

Briefly discuss what other models can you use for analysing the consumer profile on refined oil in this illustration

.....

.....

.....

.....

.....

Questionnaire

A survey is being conducted on refined oil, would you please take a few minutes of your time to answer the following questions? Thank you.

1 i) Please list the names of the brands of refined oil which you are aware of Indicate against each brand that you have mentioned

ii) Whether it is groundnut oil/Corn oil/Any other

iii) Whether you have used one or more of the brands (You may tick the brands you have used)

.....	.....
.....	.....
.....	.....
.....	.....

2 How have you come to know these brands of refined oil you have mentioned? (Source may be indicated)

3 What cooking medium do you use for cooking?

Vanaspati	Refined	Raw Oil	Any other (Mention the Medium)

Name of the Brand

Monthly Consumption (Kg.)

Refined Oil Consumers

a) Are you using refined oil on the advice of Doctor?

Yes.....No. .....

b) Indicate :
Name of the brand which you are using at present .....

.....
Monthly consumption (Kg.) .....

c) What size do you buy everytime? ——— ——— ——— ———
1 kg tin 2 kg tin 4 kg tin Bulk

d) If you are buying in loose, how much do you buy everytime?

e) Please rank the following factors in the order of most important to least important (you may assign ranks 1,2,3, etc. 1 means most important 2 next most important etc....)

Odourless (No smell)

Retains taste of food cooked

No cholesterol

Purity

Good for heart

Priced reasonably

Colour

f) For any two brands of refined oil which you have used, rate the following attributes on a 0 to 10 scale. If you feel any attribute according to you gets a grade of —

Very good	assign	10 marks
Good	assign	8 marks
Somewhat good	assign	4 marks
Not so good	assign	2 marks
Bad/absent	assign	0 marks

	Brand 1 (Mention brand)	Brand 2 (Mention brand)
Odourless (No smell)		
Taste of food cooked		
Purity		
Good for heart		
No smoke while cooking		
No cholesterol		

5 Demographic Data

a) Family size (indicate total members in your family)

b) Income/Month (Rs.) Upto 2000 2000-3000 above 3000

c) Age group of all members (Put the No. of members against each age group)

Less than 10	10-20	20-30	30-40	40 & above

15.3 ILLUSTRATION ON 'DEMAND FORECASTING'

Demand forecasting is perhaps the most frequently used tool of Marketing Research in India. All organisations want to forecast the industry demand first and then applying their market share as multiplier arrive at their sales forecast. In India both long range and short range demand forecasting studies are being undertaken from time to time. However analytical models like regression and time series analysis are not always used. Simple growth rate coupled with past average and extrapolation using graphs are the most popular methods of making forecast in Indian organisations.

In this illustration we propose to discuss the demand forecasting of Light Commercial Vehicles (LCVs) in India. With the advent of new generation of LCVs introduced by TELCO, DCM and Eicher, LCVs forecast assumes greater importance. On behalf of 'Ministry of Industry', A.F. Ferguson & Associates did a study on long term demand projections for automotive vehicles. We will discuss the regression model attempted by A.F. Ferguson and Associated and another independent study based on time series model.

Selected model of Fergusons

LCV Sales = – 136999 + 19657.2 (Δ POP I) + 2.62686 (Δ SRL) + 194 (Δ CCE)

where Δ POP I = Change in the population of Class I towns
SRL = Change in surfaced road length
CCE = Change in the amount of Credit extended to road transport operations by public sector banks.

Based on past data LCV sales is treated as a function of these three variables and a multiple regression model was attempted.

$R^2 = 0.99$
F (test value) = 163.5; (equation significant statistically at 5% level)

Sensitivity analysis

5% increase in	**Corresponding change in LCV Sales**
Population of Class I towns	21.6% (increase)
Surfaced road length	7.9% (increase)
Credit extended to the road transport operators	2.3% (increase)

Based on this equation, LCV sales in 1989-90 was forecast after substituting the forecast values of the independent variables in the equation. The LCV forecast arrived at was 81000 units. in 1989-90. If the current level of LCV production is any indication (around 48000 units), it is unlikely that the demand will reach anywhere near 81000 units. The model has not been able to predict the LCV demand with any accuracy. This is primarily because of the inherent problem associated with any regression model. If you make any mistake in forecasting the independent variables, your demand will be wrongly forecasted. Also for long term projections, the regression model will work only in certain ranges of value of independent variables. Therefore it is imperative that for continued accuracy of the model, actual values of the independent variables as and when become available should be fed back in to the system.

Time Series Model

In this model, we have taken past data on LCV sales for the period 1976 to 1985 (10 years data) and treated time as independent variable. After the model is fitted, we have made forecast for 1986, 87 and 88. Since actuals are available for 1986, 87 and 88, we can compare the forecast and actual sales to measure the power of the model to forecast. Here also we use the simple regression of LCV sales as function of time.

Period	Time	LCV Sales
1976	1	8000
77	2	7000
78	3	10000
79	4	12000
80	5	17000
81	6	20000
82	7	25000
83	8	27000
84	9	33000
85	10	35000

Results of the analysis (based on computer output)

Variable	Mean	Variance	Standard Deviation
X (Time)	5.5	9.17	3.03
Y (LCV Sales)	19400	1.056 F + 08	10276.19

Fitted Linear regression equation

$Y = 1000 + (3345.45) \times$ (Significant at 5% level)

Correlation coefficient = 0.99
Coefficient of Determination = 0.97
t ratio = 16.52
Degrees of freedom = 8
Std. deviation of errors = 1838.97

Period	LCV Sales	Estimated
1976	8000	4345
77	7000	7691
78	10000	11036
79	12000	14382
80	17000	17727
81	20000	21073
82	25000	24418
83	27000	27764
84	33000	31109
85	35000	34455

	Forecast using the model	actuals available
1986	37800	38000
1987	41145	41000
1988	44491	46000

The acid test for any forecasting model is its ability to forecast values closer to actuals there by minimising errors. The time series forecast model has given a much better reliability than the multiple regression model of A.F. Ferguson. You must remember one thing that if in any subsequent year, there is a sudden and steep upward swing in the LCV sales, the time series model may fail totally. Also there are situations where the multiple regression forecast may give better reliability. It is the responsibility of the market researcher to choose the appropriate model by doing a thorough exploratory study.

Activity 2

Suggest how you would improve the above forecasting models used in this illustration.

.....

.....

.....

.....

15.4 ILLUSTRATION ON SALES PROMOTION CAMPAIGN

It is not necessary that one should always use analytical tools of marketing research which require a great deal of sophistication. Even analysis involving simple arithmetic and descriptive statistics can bring the essence and lead to effective decisions. What is important is conceptualisation of the problem with clarity of thought. The present illustration will bring out this fact.

Sales promotion consists of a wide variety of tactical promotion tools of a short term incentive nature, designed to stimulate strong target market response. Among the more popular ones are premiums, couponing, contests, incentives and deals.

Sales promotion in the marketing mix assumes greater importance due to emergence of new product, growth of self service retailing, and heightened competition.

Evaluation of a sales promotion programme depends upon the nature of the product viz., consumer or industrial, objectives set forth like greater market share or sustaining the market share if the product is a matured one, nature of the market and competitive conditions. Evaluation may involve comparing the targets with the actual where the targets aimed at are based on the promotion scheme, or comparing the market share before the sales promotion, immediately after the promotion and say 6 months after the promotion.

In this assignment, we will evaluate a sales promotion programme (SPP) of a consumer product launched by Richardson Hindustan Limited (RHL now P&G India Ltd.) during winter 1981-82. The product is 'VICKS VAPORUB'.

Objectives

The main objectives of the sales promotion programme campaigns were :

1 To achieve large scale jar upgrading with a view to sustaining the overall market share,

2 To encourage consumers to buy VICKS VAPORUB.

3 To sample inhaler to new users and hence develop an independent inhaler franchise, and

4 To sample Vicks cough drops.

The SPP Offers

The promotion tool had three basic offers each having different pack/incentive combinations. Only one of them was applicable to a specific market, which had been decided, based on critical characteristics of the market vis-a-vis Vaporub.

The Combinations were :

	With purchase of	Consumer gets a strip of	States
1	Vaporub 19 Gm	2 vicks cough drops	Maharashtra, MP, West Bengal
	Vaporub 35 Gm	4 vicks cough drops	Assam, Orissa, Bihar
	Vaporub 60 Gm	1 Inhaler Free	Punjab/Haryana, UP and Nepal
2	Vaporub 36 Gm	Soap Dish Free	AP, Kerala, Rajasthan
	Vaporub 60 Gm	Soap Dish Free	
3	Vaporub 60 Gm	1 Inhaler Free	Bombay, Calcutta, Madras Delhi and Gujarat.

Why there offers of SPP of different types

Offer 1 : Aimed at upgrading the consumers of these states to Vaporub 19 Gm, 35 Gm and 60 Gm as the contribution of Vaporub 5 Gm, in terms of share was very high in these states.

Offer 2 : These were salient vaporub 19 Gm markets and hence the consumers of these states were to be upgraded to 35 Gm and 60 Gm.

Offer 3 : All these were well developed jar markets and/or metro markets with high purchasing power, hence focus was on upgrading consumers to the largest pack and achieve collateral sampling of 'Inhaler'

Timing of SPP

December Cycle	January Cycle
Madras, Tamilnadu, Karnataka Kerala, Andhra Pradesh	Bombay, Maharashtra, Gujarat, MP, Delhi, Punjab/Haryana, Rajasthan, UP Calcutta, West Bengal, Assam, Orissa, Bihar, Nepal.

Inputs for SPP

To extend full support to the campaign, considerable inputs were given :

1 Attractive posters mentioning the gift offer, had been allocated to each town in liberal quantities. For each offer, appropriate posters were printed.

2 In Bombay/Delhi, special TV spots were arranged on the air during prime time to indicate the offer.

3 There were press ads in Gujarat, Assam and Karnataka, to create awareness about the offer.

4 Vaporub cycle coincided with the SPP to enable RHL to get maximum impact through merchandising.

5 In addition to all these, the regular vaporub winter film was on full thrust.

Packaging of Stocks/Incentives

With vaporub 19 Gm, 2 Vicks cough drops were given free. The vicks cough drops were packed in the foil with message 'free sample — Not for Sale' printed across it. These foils were packed alongwith Vaporub 19 Gm jars, within a shrink wrap tray. Similarly 2 foils of 2s (4 Vicks cough drops) were offered free with Vaporub 35 Gm Jars, and packed in the same manner. The inhaler given free with Vaporub 60 Gm was taped on the jar.

On the cap of each SPP Jar, the free offer message had been printed, so that the consumers were aware of what was given free, when they bought it.

Soap dish incentives for Andhra Pradesh, Kerala, and Rajasthan were packed separately like in the normal merchandising SPP Scheme.

The cornerstone for success of this campaign was its ability to attract by superb display at the outlet any visitor or even somebody passing by.

Our Evaluation of the sales promotion scheme launched by Richardson Hindustan Limited now Proctar & Gamble India Ltd. (Vicks Vaporub)

We would consider the scheme to be a very creative one as well as a very effective sales promotion programme for the following reasons:

1 For the first time Cough drops/Inhaler were given as incentive at National Level. This certainly ensured a large scale jar upgrading and sustainability of existing market share at any rate. It may be pointed out here, the offtake of vaporub tied up with cold products (Inhaler) went up by a spectacular and staggering figure of 200% in test markets during the Vaporub 60 Gm Inhaler offer.

2 It is important to note that during the offer period of the above incentive, inhaler offtake was not at all affected. On the contrary the offtake of inhaler picked up in the subsequent months and hence paved the way for independent consumer franchise which was one of the objectives set forth in the sales promotion programme.

3 In case of Vaporub 60 Gm, consumer got a fantastic 30% worth of purchase as a free offer.

4 This offer had been supported with fantastic point of sales and media in selected areas.

Quantitative Evaluation

Background

The Company (RHL now P&G India Ltd.) is highly professionalised in its marketing setup and uses the modern tools of evaluation. It does compare the estimated and actual figures in respect of market share, offtakes, distribution and stock turnover ratio before and after any promotion. It also gets a feedback from external source like ORG, pioneer in retail audit. In view of the fact, the latest information in the above areas could not be parted with by the company, we give below certain quantitative data which highlights the fact that RHL is a market leader in this product, head and shoulder above its main competitors namely 'Amrutanjan' and 'Rubex'.

Before giving the quantitative evaluation of this particular sales promotion scheme, it may be pointed out in the passing that the company had been able to sustain its market share consistently only due to its systematic and creative sales promotional activities which are of continuous nature. A decline in sales is followed by a sales promotion programme with creative incentive offer to induce the consumer to buy more. The data with regard to sustained market share, distribution and stock turnover ratio of RHL (now P&G India Ltd.) in comparison with Amrutanjan and Rubex is given below (sufficient gap in time was given after the campaign was launched).

i) **% Market Share in Urban Market**

Region	RHL	Amrutanjan	Rubex	Others
All India	62	35	2	1
North	84	14	1	1
East	73	22	1	4
West	69	26	3	2
South	46	53	1	–

It may be seen from above that RHL was not only sustaining its all India market share of 62% and was not only the market leader in every region except south, but also it was catching up fast in South by getting 46% of the market as against 53% by Amrutanjan which is situated in the heart of Madras City. This commendable achievement was entirely due to its sales promotion programme and incentive offers incorporated in the scheme as discussed earlier.

ii) With regard to the number of retail outlets (urban) RHL had 179 thousand outlets at the National Urban Market as against 121 thousand outlets of Amrutanjan and 24 thousand outlets of Rubex.

iii) If we define the stock turnover ratio index as

$$\frac{\text{STR (RHL)}}{\text{STR (Competitors)}} \times 100,$$ it worked out to 146.

The interpretation is that the Stock Turnover Ratio of Richardson Hindustan Ltd., was 146 times as that of competitors or RHL was able to turn out a sales volume of 146% of its competitors.

To sum up, the present sales promotion programme of Richardson Hindustan Limited (now P&G India Ltd.) was effectively in every sense of market response and the company would continue to be a market leader in this line of Rubs and Balms because of the creativity involved in the scheme. The present sales promotion programme which had been in vogue for some time, achieved the major objective of upgradation of jars and sustained market share as well as an independent franchise for inhaler.

Activity 3

Can you use analysis of variance tool in this example? If so, what kind of field experiment would you conduct ?

.....

15.5 OTHER POTENTIAL APPLICATIONS

The illustrations we have covered so far on the application of Marketing Research in India are the most popular ones in the sense that they are being frequently used by the organisations and research firms. However there is still a vast reservoir of untapped potential that exists in terms of marketing research applications. We shall list below the potential areas of applications.

1 **New product launching and test marketing in India,** require marketing research tools to aid decision making. Hindustan Lever did a classic test marketing before introducing 'LIRIL' soap at all India level. The organisation tested this product in two typical cities—Hyderabad in South and Lucknow in the North. Since this type of study involves experimental research, analysis of variance is the appropriate tool to use.

2 **Measuring the advertising and sales promotion effectiveness** involves good field experiment. Therefore if the objective is to measure the improvement in sales contributed by the campaign, it is important to use analytical tools like analysis of variance and paired t test. Currently the practice is to use simple arithmetic to calculate market share before and after the campaign. It is important that one should test the significance statistically.

3 **Media planning and determination of media mix** is being handled exhaustively by advertising firms and many private organisations. Stimulation techniques and linear programming models are being attempted to decide optimal media mix. Because of the conflicting goals encountered in reality, "Goal Programming model" is perhaps potentially the answer to the problem of media mix.

4 **Communicating research** offers good scope for marketing research to find out whether the communication is effective. For example one may be interested in finding out whether the message conveyed through an advertisement is perceived properly by the target audience. Suppose through a questionnaire, if you can collect data on the understanding aspect of the consumer on some suitable scale. you can use discriminant analysis to classify the consumer into two groups — one which has understood the campaign — another which has not understood the campaign.

5 **Plethora of marketing data** bewilders marketing managers and acts as a bottle-neck for extracting relevant information for decision making. Factor analysis could be used to collapse the data and reduce them into fewer factors for analysis. Stepwise regression can also be used for this purpose.

It may be mentioned here that the above applications are only indicative and certainly not an exhaustive list. Further scope exists which should be fully exploited. However the above mentioned applications are the typical areas of potential applications.

15.6 SUMMARY

We have then covered three typical illustrations on 'consumer research', 'demand forecasting', and 'sales promotion campaign' in the Indian Scene. 'Consumer Research' illustration on refined oil has focused on problems concerning consumer buying behaviour and the strategies needed to take appropriate actions when called for. 'Demand Forecasting' illustration attempts to forecast realistically the demand for Light Commercial Vehicles (LCV). With actual example, the strengths and limitations of the multiple regression model and simple linear regression based on

time series data have been highlighted. 'Sales Promotion Campaign' illustration has brought out the imortance of descriptive aspect of communication research and using simple calculation in the context of evaluating the campaign. The illustration is on the sales promotion campaign launched by 'RICHARDSON HINDUSTAN' for 'VICKS VAPORUB' We have then moved on to 'other potential applications' of marketing research in India. We have listed some typical areas of application involving marketing research where the potential is vast. They include — new product launching and test marketing — measuring the advertising and sales promotion effectiveness, — media planning and determination of media mix, communication research, and 'collapsing data' for meaningful interpretation. It has also been mentioned that this list is only indicative and could include other areas as well which will require marketing research help.

15.7 SELF-ASSESSMENT EXERCISES

1. A well established consumer oriented company is engaged in the manufacture of detergent cakes and powder. The Company wants to do in depth study on the consumer profile of detergent cake to readjust its brand positioning. How would you go about doing this study?
2. Suppose you are the marketing manager of a newly estabiished electronic industry interested in manufacturing Televisions (TVs). How would you go about assessing the present demand for TVs in the country and the forecast for the next five years?
3. Suggest your own methods for overcoming the resistance to using market research if you are employed as marketing head of a large manufacturing organisation.

15.8 FURTHER READINGS

1) F.E. Brown, *Marketing Research* Addison-Wesley Publishing Company,

2) Paul E. Green and Donald S. Tull, *Research for Marketing Decisions,* Prentice Hall of India.