LOVELY PROFESSIONAL UNIVERSITY

QUANTITATIVE TECHNIQUES-II

Edited By

Neha Tikoo

PPrinted by
EXCEL BOOKS PRIVATE LIMITED
A-45, Naraina, Phase-I,
New Delhi-110028
for
Lovely Professional University
Phagwara

SYLLABUS

Quantitative Techniques-II

Objectives: To familiarize the students with different statistical techniques useful in conducting business research and then applying the same in their business strategies.

Sr. No.	Description
1.	**Quantitative techniques for managers :** quantitative decision making & its overview, An introduction to research: meaning, definition and objectives, Goals, Strategy, Tactics, Internal and External Research Suppliers, Research Method Concept, Constructs, Definitions, Variables, Propositions and Hypotheses research process
2.	**Research problem:** selection of problem, understanding problem, necessity of defined problem, Pilot Testing, Data Collection, Analysis and Interpretation, Reporting the Results. Review of literature in research
3.	**Research design:** meaning, types – descriptive, diagnostic, exploratory and experimental
4.	**Sources and methods of data collection:** primary and secondary sources, data collection methods, Questionnaire designing: construction, types, developing a good questionnaire, mailed questionnaire and schedule
5.	**Sampling design and techniques, Scaling techniques:** meaning and types, sampling distribution, Data processing operations: editing, coding, classification, tabulation,
6.	**Partial Correlation:** zero order, first order, second order Multiple Correlation, coefficient of Multiple correlation
7.	**Multiple Regression and Correlation Analysis:** Least square regression plane, linear Multiple regression analysis, Coefficient of Multiple Determination
8.	**Hypothesis Testing:** Statistical significance, the logic of hypothesis testing, statistical testing procedure, p-values.
9.	**Test of significance:** Types of tests, z-test, t-test, chi-square test, ANOVA
10.	Factor Analysis, Cluster Analysis and Conjoint Analysis

CONTENT

Unit 1:	Quantitative Techniques for Managers *Neha Tikoo, Lovely Professional University*	1
Unit 2:	Introduction to Research *Neha Tikoo, Lovely Professional University*	19
Unit 3:	Language of Research *Neha Tikoo, Lovely Professional University*	41
Unit 4:	Research Problem *Neha Tikoo, Lovely Professional University*	53
Unit 5:	Review of Literature in Research *Neha Tikoo, Lovely Professional University*	74
Unit 6:	Research Design *Neha Tikoo, Lovely Professional University*	80
Unit 7:	Sources and Methods of Data Collection *Neha Tikoo, Lovely Professional University*	107
Unit 8;	Sampling and Sampling Distribution *Dilfraz Singh, Lovely Professional University*	142
Unit 9:	Attitude Measurement and Scaling Techniques *Tanima Dutta, Lovely Professional University*	163
Unit 10:	Correlation *Dilfraz Singh, Lovely Professional University*	196
Unit 11:	Multiple Regression and Correlation Analysis *Tanima Dutta, Lovely Professional University*	217
Unit 12:	Hypothesis Testing *Dilfraz Singh, Lovely Professional University*	233
Unit 13:	Test of Significance *Tanima Dutta, Lovely Professional University*	241
Unit 14:	Multivariate Analysis *Dilfraz Singh, Lovely Professional University*	255

Unit 1: Quantitative Techniques for Managers

Notes

CONTENTS

Objectives

Introduction

1.1 Quantitative Decision-making and its Overview

1.2 Meaning of Quantitative Techniques

1.3 Statistics and Operations Research

1.3.1 Types of Statistical Data

1.3.2 Classification of Statistical Methods

1.4 Models in Operations Research

1.5 Various Statistical Techniques

1.6 Advantages of Quantitative Approach to Management

1.7 Quantitative Techniques in Business and Management

1.8 Summary

1.9 Keywords

1.10 Review Questions

1.11 Further Readings

Objectives

After studying this unit, you will be able to:

- Provide an overview of quantitative techniques;
- Know the need of using quantitative approach to managerial decisions;
- Appreciate the role of statistical methods in data analysis;
- Know the various models frequently used in operations research and the basis of their classification;
- Have a brief idea of various statistical methods;
- Discuss the areas of applications of quantitative approach in business and management.

Introduction

You may be aware of the fact that prior to the industrial revolution individual business was small and production was carried out on a very small scale mainly to cater to the local needs. The management of such business enterprises was very different from the present management of large scale business. The information needed by the decision-maker (usually the owner) to make effective decisions was much less extensive than at present. Thus, he used to make decisions based upon his past experience and intuition only. Some of the reasons for this were:

Notes

- The marketing of the product was not a problem because customers were, for the large part, personally known to the owner of the business. There was hardly any competition in the business.
- Test marketing of the product was not needed because the owner used to know the choice and requirement of the customers just by personal interaction.
- The manager (also the owner) also used to work with his workers at the shop floor. He knew all of them personally as the number were small. This reduced the need for keeping personal data.
- The progress of the work was being made daily at the work centre itself. Thus production records were not needed.
- Any facts the owner needed could be learnt direct from observation and most of what he required was known to him.

Now, in the face of increasing complexity in business and industry, intuition alone has no place in decision-making because basing a decision on intuition becomes highly questionable when the decision involves the choice among several courses of action each of which can achieve several management objectives simultaneously.

Hence there is a need for training people who can manage a system both efficiently and creatively.

1.1 Quantitative Decision-making and its Overview

Quantitative techniques have made valuable contribution towards arriving at an effective decision in various functional areas of management-marketing, finance, production and personnel. Today, these techniques are also widely used in regional planning, transportation, public health, communication, military, agriculture, etc.

Quantitative techniques are being used extensively as an aid in business decision making due to following reasons:

- Complexity of today's managerial activities which involve constant analysis of existing situation, setting objectives, seeking alternatives, implementing, coordinating, controlling and evaluating the decision made.
- Availability of different types of tools for quantitative analysis of complex managerial problems.
- Availability of high speed computers to apply quantitative techniques (or models) to real life problems in all types of organizations such as business, industry, military, health, and so on. Computers have played an important role in arriving at the optimal solution of complex managerial problems both in terms of time and cost.

In spite of these reasons, the quantitative approach, however, does not totally eliminate the scope of qualitative or judgment ability of the decision-maker. Of course, these techniques complement the experience and knowledge of decision maker in decision-making.

Self Assessment

Fill in the blanks:

1. ……are also widely used in regional planning, transportation, public health, communication, military, agriculture, etc.

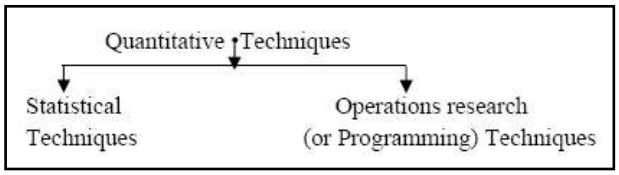
Notes

2. The quantitative approach does not totally eliminate the scope of …….ability of the decision-maker.

Task Explain with the help of example some of the important Quantitative Techniques used in modern business and in industrial unit.

1.2 Meaning of Quantitative Techniques

Quantitative techniques refer to the group of statistical, and operations research (or programming) techniques as shown in the following chart.



The quantitative approach in decision-making requires that, problems be defined, analysed and solved in a conscious, rational, systematic and scientific manner based on data, facts, information, and logic and not on mere whims and guesses. In other words, quantitative techniques (tools or methods) provide the decision-maker a scientific method based on quantitative data in identifying a course of action among the given list of courses of action to achieve the optimal value of the predetermined objective or goal. One common characteristic of all types of quantitative techniques is that numbers, symbols or mathematical formulae (or expressions) are used to represent the models of reality.

1.3 Statistics and Operations Research

The word statistics can be uses, in a number of ways. Commonly it is described in two senses namely:

1. ***Plural Sense (Statistical Data):*** The plural sense of statistics means some sort of **statistical data.** When it means statistical data, it refers to numerical description of quantitative aspects of things; These descriptions may take the form of counts or measurements.

Example: Statistics of students of a college include count of the number of students, and separate counts of number of various kinds as such, male and females, married and unmarried, or undergraduates and post-graduates. They may also include such measurements as their heights and weights.

2. ***Singular Sense (Statistical Methods):*** The large volume of numerical information (or data) gives rise to the need for systematic methods which can be used to **collect, organise** or **classify, present, analyse** and **interpret** the information effectively for the purpose of making wise decisions. Statistical methods include all those devices of analysis and synthesis by means of which statistical data are systematically collected and used to explain or describe a given phenomena. The above mentioned five functions of statistical methods are also called **phases** of a statistical investigation.

   Methods used in analysing the presented data are numerous and contain simple to sophisticated mathematical techniques.

Notes

As an illustration, let us suppose that we are interested in knowing the income level of the people living in a certain city. For this we may adopt the following procedures:

(i) *Data collection:* The following data is required for the given purpose:

(a) Population of the city

(b) Number of individuals who are getting income

(c) Daily income of each earning individual

(ii) *Organise (or condense) the data:* The data so obtained should now be organised in different income groups. This will reduce the bulk of the data.

(iii) *Presentation:* The organised data may now be presented by means of various types of graphs or other visual aids. Data presented in an orderly manner facilitates statistical analysis.

(iv) *Analysis:* On the basis of systematic presentation (tabular form or graphical form), determine the average income of an individual and extent of disparities that exist. This information will help to get an understanding of the phenomenon (i.e. income of 'individuals).

(v) *Interpretation:* All the above steps may now lead to drawing conclusions which will aid in decision-making-a policy decision for improvement of the existing situation.

Characteristics of Data

It is probably more common to refer to data in quantitative form as statistical data. But not all numerical data is statistical. In order that numerical description may be called statistics they must possess the following characteristics:

- ***They must be aggregate of facts***, for example, single unconnected figures cannot be- used to study the characteristics of the phenomenon.
- ***They should be affected to a marked extent by multiplicity of causes***, for example, in social services the observations recorded are affected by a number of factors (controllable and uncontrollable).
- ***They must be enumerated or estimated according to reasonable standard of accuracy***, for example, in the measurement of height one may measure correct up to 0.01 of a cm; the quality of the product is estimated by certain tests on small samples drawn from a big lot of products.
- ***They must have been collected in a systematic manner for a pre-determined purpose***. Facts collected in a haphazard manner and without a complete awareness of the object, will be confusing and cannot be made the basis of valid conclusions.

Example: Collected data on price serve no purpose unless one knows whether he wants to collect data on wholesale or retail prices and what are the relevant commodities in view.

- ***They must be placed in relation to each other.*** That is, data collected should be comparable; otherwise these cannot be placed in relation to each other, e.g. statistics on the yield of crop and quality of soil are related but these yields cannot have any relation with the statistics on the health of the people.
- ***They must be numerically expressed.*** That is, any facts to be called statistics must be numerically or quantitatively expressed. Qualitative characteristics such as beauty, intelligence, etc. cannot be included in statistics unless they are quantified.

Notes

1.3.1 Types of Statistical Data

An effective managerial decision concerning a problem on hand depends on the availability and reliability of statistical data. Statistical data can be broadly grouped into two categories:

- Secondary (or published) data
- Primary (or unpublished) data

The secondary data are those which have already been collected by another organisation and are available in the published form. You must first check whether any such data is available on the subject matter of interest and make use of it, since it will save considerable time and money. But the data must be scrutinized properly since it was originally collected perhaps for another purpose. The data must also be checked for reliability, relevance and accuracy.

A great deal of data is regularly collected and disseminated by international bodies such as: World Bank, Asian Development Bank, International Labour Organisation, Secretariat of United Nations, etc., Government and its many agencies: Reserve Bank of India, Census Commission, Ministries-Ministry of Economic Affairs, Commerce Ministry; Private Research Organizations, Trade Associations, etc.

When secondary data is not available or it is not reliable, you would need to collect original data to suit your objectives. Original data collected specifically for a current research are known as primary data. Primary data can be collected from customers, retailers, distributors, manufacturers or other information sources. Primary data may be collected through any of the three methods: observation, survey, and experimentation.

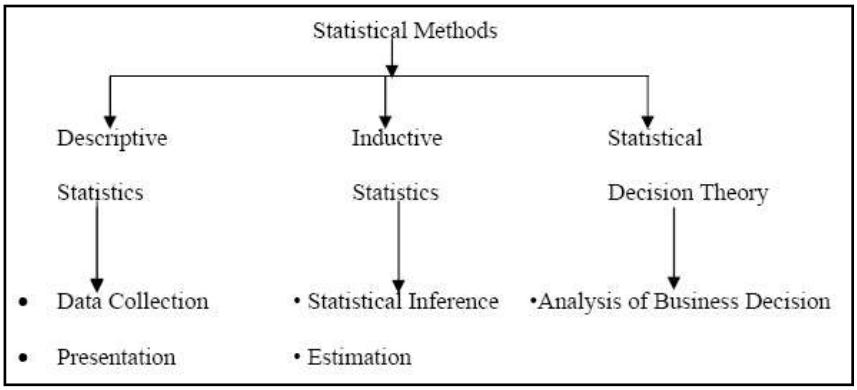
Data are also classified as micro and macro. Micro data relate to a particular unit or region whereas macro data relate to the entire industry, region or economy.

Operations Research

You would recall that in Operations Research a mathematical model to represent the situation under study is constructed. This helps in two ways. Either to predict the performance of the system under certain controls or to determine the action or control needed to optimise performance.

1.3.2 Classification of Statistical Methods

By now you may have realised that effective decisions have to be based upon realistic data. The field of statistics provides the methods for collecting, presenting and meaningfully interpreting the given data. Statistical Methods broadly fall into three categories as shown in the following chart.



Notes

Descriptive Statistics

There are statistical methods which are used for re-arranging, grouping and summarising sets of data to obtain better information of facts and thereby better description of the situation that can be made.

Example: Changes in the price-index. Yield by wheat etc. are frequently illustrated using different types of charts and graphs.

These devices summarise large quantitatives of numerical data for easy understanding. Various types of averages can also reduce a large mass of data to a single descriptive number. The descriptive statistics include the methods of collection and presentation of data, measure of Central tendency and dispersion, trends, 'index numbers, etc.

Inductive Statistics

It is concerned with the development of some criteria which can be used to derive information about the nature of the members of entire groups (also called population or universe) from the nature of the small portion (also called sample) of the given group. The specific values of the population members are called 'parameters' and that of sample are called 'statistics'. Thus, inductive statistics is concerned with estimating population parameters from the sample statistics and deriving a statistical inference.

Samples are drawn instead of a complete enumeration for the following reasons:

- The number of units in the population may not be known.
- The population units may be too many in number and/or widely dispersed. Thus complete enumeration is extremely time consuming and at the end of a full enumeration so much time is lost that the data becomes obsolete by that time.
- It may be too expensive to include each population item.

Inductive statistics, includes the methods like: probability and probability distributions; sampling and sampling distributions; various methods of testing hypothesis; correlation, regression, factor analysis; time series analysis.

Statistical Decision Theory

Statistical decision theory deals with analysing complex business problems with alternative courses of action (or strategies) and possible consequences. Basically, it is to provide more concrete information concerning these consequences, so that best course of action can be identified from alternative courses of action.

Statistical decision theory relies heavily not only upon the nature of the problem on hand, but also upon the decision environment. Basically there are four different states of decision environment as given below:

State of decision	Consequences
Certainty	Deterministic
Risk	Probabilistic
Uncertainty	Unknown
Conflict	Influenced by an opponent

Notes

Since statistical decision theory also uses probabilities (subjective or prior) in analysis, therefore it is also called a subjectivist approach. It is also known as Bayesian approach because Baye's theorem is used to revise prior probabilities in the light of additional information.

Self Assessment

Fill in the blanks:

3. ……include all those devices of analysis and synthesis by means of which statistical data are systematically collected and used to explain or describe a given phenomena.
4. The ……are those which have already been collected by another organisation and are available in the published form.
5. Statistical decision theory relies heavily not only upon the nature of the problem on hand, but also upon the ……

1.4 Models in Operations Research

In this Section we are presenting several classifications of OR models so that you should know more about the role of models in decision-making:

Purpose

A Model is the representation of a system which, in turn, represents a specific part of reality (an object of interest or subject of inquiry in real life). The means of representing a system may be physical, graphic, schematic, analogy, mathematical, symbolic or a combination of these. Through all these means, an attempt is made to abstract the essence of reality, which in turn, is quite helpful to **describe, explain** and predict the behaviour of the system Thus, depending upon the purpose, the stage at which the model is developed, models can be classified into four categories.

1. ***Descriptive model:*** Such Models are used to describe the behaviour of a system based on certain information.

Example: A model can be built to describe the behaviour of demand for an inventory item for a stated period, by keeping the record of various demand levels and their respective frequencies.

A descriptive model is used to display the problem situation more vividly including the alternative choices to enable the decision-maker to evaluate results of each alternative choice. However, such model does not select the best alternative.

2. ***Explanatory model:*** Such models are used to explain the behaviour of a system by establishing relationships between its various components.

Example: A model can be built to explain variations in productivity by establishing relationships among factors such as wages, promotion policy, education levels, etc.

3. ***Predictive model:*** Such models are used to predict the status of a system in the near future based on data.

Example: A model can be built to predict stock prices (within an industry group), for given any level of earnings per share.

Notes

4 ***Prescriptive (or normative) model:*** A prescriptive model is one which provides the norms for the comparison of alternative solutions which result in the selection of the best alternative (the most preferred course of action).

Example: Allocation models.

Degree of Abstraction

The following chart shows the classification of models according to the degree of abstraction:

Model	Degree of Abstraction
• Physical	Least Abstract
• Graphic	↑
• Schematic	
• Analog	↓
• Mathematical	Most Abstract

Any three-dimensional model that looks like the real thing but is either reduced in size or scaled up, is a physical (conic) model. These models include city planning maps, plant layout charts, plastic model of airplane, body parts, etc. These models are easy to observe, build and describe, but cannot be manipulated and used for prediction.

An organisation chart showing responsibility relationships is an example of graphic model. A flow chart (or diagram) depicting the sequence of activities during the complete processing of a product is an example of schematic model. Another example of schematic model is the Computer programme where main features of the programme are represented by a schematic description of steps.

Analog models are closely associated with iconic models. However, they are not replicas of problem situations. Rather they are small physical systems that have similar characteristics and work like an object or system it represents.

Example: Children's toys, model rail-roads, etc.

These models might not allow direct handling or manipulation.

Mathematical (or symbolic) models represent the systems (or reality) by using mathematical symbols and relationships. These are very precise, most abstract and can be manipulated by using laws of mathematics. The input-output model of national economy involving several objectives, constraints, inputs and inter-linkages between them is an example of representing a complex system with the help of a set of equations.

Degree of Certainty

Models can also be classified according to the degree of assumed certainty. Under this classification models are divided into deterministic versus probabilistic models.

Models in which selection of each course of action (or strategy) results in unique and known pay-off or consequence are called deterministic models.

Example: Linear programming, transportation and assignment models.

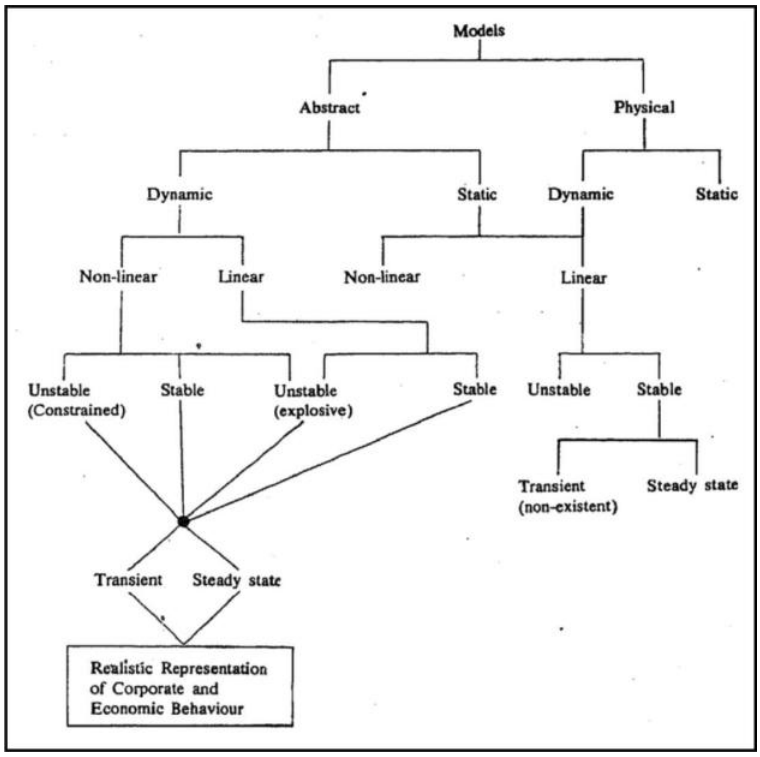
Notes

Situations in which each course of action (or strategy) can result in more than one pay-offs or consequences are called probabilistic models. Since in such models the concept of probability is used, therefore the pay-off or consequences due to a managerial action cannot be predicted with certainty.

Example: Simulation models, decision theory models etc.

Specified Behaviour Characteristics

The following chart describes the classification of models based on specified behaviour characteristics. Such type of classification helps in understanding the nature and role of models in representing management and economic status of organisations.



Source: Loomba, M.P. 1978. Management-A Quantitative Perspective; Macmillan Publishing Co.: New York)

The models that are concerned with a particular set of fixed conditions and do not change in a short-term period (or planning period) are known as static models. This implies that such models are independent of time and only one decision is required for a given time period.

Notes The resources required for a product and the technology or manufacturing process do not change in short-term period.

Linear programming is the particular example of static models. On the other hand, there are certain types of problems where time factor plays an important role and admit the impact of changes over a period of time. In all such situations decision-maker has to make a sequence of optimal decisions at every decision point (i.e. variable time) regardless of what the prior decision

Notes

has been. The problem of product development in which the decision-maker has to make decisions at every decision point such as product design, test-market, full-scale production, etc. is an example of dynamic model. Dynamic programming is the particular example of dynamic model.

Linear Models are those in which each component exhibits a linear behaviour. The word 'linear' is used to describe the relationship among two or more variables which are directly proportional. For example, if our resources increase b some percentage, then it would increase the output by the same percentage.

Did u know? **What are non-linear models?**

If one or more components of a model exhibit a non-linear behaviour, then such models are called **non-linear models.**

A mathematical model of the form Z = 5 + 3 is called a linear model whereas a model of the form $Z = 5x^2 + 3xy + y^2$ is called a non-linear model.

Procedure (or Method) of Solution

The type of procedure used to derive solutions to mathematical models divides them into two categories: (i) analytical models, and (ii) simulation models.

An analytical model consists of a mathematical structure and is solved by known mathematical or analytical techniques to yield a general solution.

Example: Network models (PERT/CPM), linear programming models, game theory models, inventory control models.

A simulation model is the experimentation (Computer assisted or manual) on a mathematical structure of real-life system. It is done by inserting into the given structure specific values of decision variables under certain assumptions in order to describe and evaluate systems behaviour over a period of time.

Example: We can test the effect of different number of service counters assuming different arrival rates of customers on total cost of providing service to customers.

The following table summarises our discussion on classification of models.

Criterion Classification	Categories of OR Models
• Purpose	• Descriptive, Explanatory, Predictive, Prescriptive
• Degree of abstraction	• Physical, Graphic, Schematic, Analog, Mathematical
• Degree of Certainty	• Deterministic, Probabilistic Certainty, Risk, Uncertainty
• Specified behaviour characteristics	• Static, Dynamic, Linear Non-linear
• Procedure of solution	• Analytical, Simulation

You have read about certain standard techniques or prototype models of operations research which can be helpful to a decision-maker in solving a variety of problems.

Self Assessment

Notes

Fill in the blanks:

6. …….are those in which each component exhibits a linear behaviour.
7. A …….is the experimentation on a mathematical structure of real-life system.

1.5 Various Statistical Techniques

A brief comment on certain standard techniques of statistics which can be helpful to a decision-maker in solving problems is given below. However, each one of these techniques requires detailed studies and in our context we are merely listing these to arouse your interest.

(i) ***Measures of Central Tendency:*** Obviously for proper understanding of quantitative data, they should be classified and converted into a frequency distribution (number of times or frequency with which a particular data occurs in the given mass of data). This type of condensation of data reduces their bulk and gives a clear picture of their structure. If you want to know any specific characteristics of the given data or if frequency distribution of one set of data to be compared with another, then it is necessary that the frequency distribution itself must be summarized and condensed in such a manner that it must help us to make useful inferences about the data and also provide yardstick for comparing different sets of data. Measures of average or central tendency provide one such yardstick. Different methods of measuring central tendency provide us with different kinds of averages. The main three types of averages commonly used are:

(a) *Mean:* The mean is the common arithmetic average. It is computed by dividing the sum of the values of the observations by the number of item:: observed.

(b) *Median:* The median is that item which lies exactly half-way between the lowest and highest value when the data is arranged in an ascending or descending order. It is not affected by the value of the observation but by the number of observations. Suppose you have the data on monthly income of households in a particular area. The median value would give you that monthly income which divides the number of households into two equal parts. Fifty per cent of all the households have a monthly income above the median value and fifty per cent of households have a monthly income below the median income.

(c) *Mode:* The mode is the central value (or item) that occurs most frequently. When the data organised as a frequency distribution the mode is that category which has the maximum number of observations.

Example: A shopkeeper ordering fresh stock of shoes for the season would make use of the mode to determine the size which is most frequently sold.

The advantages of mode are that (a) it is easy to compute, (b) is not affected by extreme values in the frequency distribution, and (c) is representative if the observations are clustered at one particular value or class.

(ii) ***Measures of Dispersion:*** The measures of central tendency measure the most typical value around which most values in the distribution tend to converge. However, there are always extreme values in each distribution. These extreme values indicate the spread or the dispersion of the distribution. The measures of this spread are called 'measures of dispersion' or 'variation' or 'spread'. Measures of dispersion would tell you the number of values which are substantially different from the mean, median or mode. The commonly used measures of dispersion are range, mean deviation and standard deviation.

The data may spread around the central tendency in a symmetrical or an asymmetrical pattern.

Notes

Did u know? What is Skewness?

The measures of the direction and degree of symmetry are called measures of the **skewness.**

Another characteristic of the frequency distribution is the shape of the peak, when it is plotted on a graph paper.

Did u know? What is Kurtosis?

The measures of the peakedness are called measures of **Kurtosis.**

(iii) ***Correlation:*** Correlation coefficient measures the degree to which the change in one variable (the dependent variable) is associated with change in the other variable (independent one). For example, as a marketing manager, you would like to know if there is any relation between the amount of money you spend on advertising and the sales you achieve. Here, sales are the dependent variable and advertising budget is the independent variable. Correlation coefficient, in this case, would tell you the extent of relationship between these two variables, whether the relationship is directly proportional (i.e. increase or decrease in advertising is associated with increase or decrease in sales) or it is an inverse relationship (i.e. increasing advertising is associated with decrease in sales and vice-versa) or there is no relationship between the two variables.

Caution Correlation coefficient does not indicate a casual relationship, Sales is not a direct result of advertising alone, and there are many other factors which affect sales.

Correlation only indicates that there is some kind of association-whether it is casual or causal can be determined only after further investigation. You may find a correlation between the height of your salesmen and the sales, but obviously it is of no significance.

(iv) ***Regression Analysis:*** For determining causal relationship between two variables you may use regression analysis. Using this technique you can predict the dependent variables on the basis of the independent variables. In 1970, NCAER (National Council of Applied and Economic Research) predicted the annual stock of scooters using a regression model in which real personal disposable income and relative weighted price index of scooters were used as independent variable.

The correlation and regression analysis are suitable techniques to find relationship between two variables only. But in reality you would rarely find a one-to-one causal relationship; rather you would find that the dependent variables are affected by a number of independent variables.

Example: Sale affected by the advertising budget, the media plan, the content of the advertisements, number of salesmen, price of the product, efficiency of the distribution network and a host of other variables.

For determining causal relationship involving two or more variables, multi-variate statistical techniques are applicable. The most important of these are the **multiple regression analysis, discriminant analysis** and **factor analysis.**

(v) ***Time Series Analysis:*** A time series consists of a set of data (arranged in some desired manner) recorded either at successive points in time or over successive periods of time. The changes in such type of data from time to time are considered as the resultant of the

Notes

combined impact of a force that is constantly at work. This force has four components: (i) Editing time series data, (ii) secular trend, (iii) periodic changes, cyclical changes and seasonal variations, and (iv) irregular or random variations. With time series analysis, you can isolate and measure the separate effects of these forces on the variables. Examples of these changes can be seen, if you start measuring increase in cost of living, increase of population over a period of time, growth of agricultural food production in India over the last fifteen years, seasonal requirement of items, and impact of floods, strikes, and wars and so on.

(vii) ***Index Numbers:*** Index number is a relative number that is used to represent the net result of change in a group of related variables that has some over a period of time. Index numbers are stated in the form of percentages.

Example: If we say that the index of prices is 105, it means that prices have gone up by 5% as compared to a point of reference, called the base year. If the prices of the year 1985 are compared with those of 1975, the year 1985 would be called "given or current year" and the year 1975 would be termed as the "base year". Index numbers are also used in comparing production, sales price, volume employment, etc. changes over period of time, relative to a base.

(viii) ***Sampling and Statistical Inference:*** In many cases due to shortage of time, cost or non-availability of data, only limited part or section of the universe (or population) is examined to (i) get information about the universe as clearly and precisely as possible, and (ii) determine the reliability of the estimates. This small part or section selected: from the universe is called the sample and the process of selections such a section (or past) is called sampling.

Scheme of drawing samples from the population can be classified into two broad categories:

(a) *Random sampling schemes:* In these schemes drawing of elements from the population is random and selection of an element is made in such a way that every element has equal chance (probability) of being selected.

(b) *Non-random sampling schemes:* In these schemes, drawing of elements from the population is based on the choice or purpose of selector.

The sampling analysis through the use of various 'tests' namely Z-normal distribution, student's 't' distribution; F-distribution and x^2-distribution make possible to derive inferences about population parameters with specified level of significance and given degree of freedom. You will read about a number of tests in this block to derive inference about population parameters.

Self Assessment

State whether true or false:

8. Regression measures the degree to which the change in one variable (the dependent variable) is associated with change in the other variable (independent one).

9. Index number is a relative number that is used to represent the net result of change in a group of related variables that has some over a period of time.

1.6 Advantages of Quantitative Approach to Management

Executives at all levels in business and industry come across the problem of making decision at every stage in their day-to-day activities. Quantitative techniques provide the executive with scientific basis for **decision-making** and **enhance his ability** to matte long-range plans and to

Notes

solve every day problems of running a business and industry with greater efficiency and confidence.

Let us now also look at some of the advantages of the study of statistics:

1. ***Definiteness:*** The study of statistics helps us in presenting general statements in a precise and a definite form. Statements of facts conveyed numerically are more precise and convincing than those stated qualitatively.

Example: The statement that "literacy rate as per 1981 census was 36% compared to 29% for 1971 census" is more convincing than stating simply that "literacy in our country has increased".

2. ***Condensation:*** The new data is often unwieldy and complex. The purpose of statistical methods is to simplify large mass of data and to present meaningful information from them.

Example: It is difficult to form a precise idea about the income position of the people of India from the data of individual income in the country. The data will be easy to understand and more precisely if it can be expressed in the form of per capita income.

3. ***Comparison:*** According to Bodding, the object of statistics is to enable comparisons between past and present results with a view to ascertaining the reasons for change which have taken place and the effect of such changes in the future. Thus, if one wants to appreciate the significance of figures, then he must compare them with other of the same kind.

Example: The statement "per capita income has increased considerably" shall not be meaningful unless some comparison of figures of past is made. This will help in drawing conclusions as to whether the standard of living of people of India is improving.

4. ***Formulation of policies:*** Statistics provides the basic material for framing policies not only in business but in other fields also.

Example: Data on birth and mortality rate not only help is assessing future growth in population but also provide necessary data for framing a scheme of family planning.

5. ***Formulating and testing hypothesis:*** Statistical methods are useful in formulating and testing hypothesis or assumption or statement and to develop new theories.

Example: The hypothesis: "whether a student has benefited from a particular media of instruction", can be tested by using appropriate statistical method.

6. ***Prediction:*** For framing suitable policies or plans, and then for implementation it is necessary to have the knowledge of future trends. Statistical methods are highly useful for forecasting future events.

Example: For a businessman to decide how many units of an item should be produced in the current year, it is necessary for him to analyse the sales data of the past years.

1.7 Quantitative Techniques in Business and Management

Notes

Some of the areas where statistics can be used are as follows:

Management

(i) ***Marketing:***

(a) Analysis of marketing research information

(b) Statistical records for building and maintaining an extensive market

(c) Sales forecasting

(ii) ***Production:***

(a) Production Planning, control and analysis

(b) Evaluation of machine performance

(c) Quality control requirements

(d) Inventory control measures

(iii) ***Finance, Accounting and Investment:***

(a) Financial forecast, budget preparation

(b) Financial investment decisions

(c) Selection of securities

(d) Auditing function

(e) Credit policies, credit risk and delinquent accounts

(iv) ***Personnel:***

(a) Labour turnover rate

(b) Employment trends

(c) Performance appraisal

(d) Wage rates and incentive plans

Economics

- Measurement of gross national product and input output analysis
- Determination of business cycle, long-term growth and seasonal fluctuations
- Comparison of market prices, cost and profits of individual firms
- Analysis of population, land economics and economic geography
- Operational studies of public utilities
- Formulation of appropriate economic policies and evaluation of their effect

Research and Development

- Development of new product lines
- Optimal use of resources
- Evaluation of existing products

Notes

Natural Science

- Diagnosing the disease based on data like temperature, pulse rate, blood pressure etc.
- Judging the efficacy of a particular drug for curing a certain disease
- Study of plant life

Task Think of any major decision you made recently. Recall the steps taken by you to arrive at the final decision. Prepare a list of those steps.

Self Assessment

State whether true or false:

10. Statements of facts conveyed numerically are more precise and convincing than those stated qualitatively.
11. Statistical methods are useful in formulating and testing hypothesis or assumption or statement and to develop new theories.
12. Statistical methods are not useful for forecasting future events.

1.8 Summary

- Quantitative techniques refer to the group of statistical, and operations research (or programming) techniques.
- The word statistics can be uses, in a number of ways. Commonly it is described in two senses namely: Plural Sense (Statistical Data) and Singular Sense (Statistical Methods).
- The field of statistics provides the methods for collecting, presenting and meaningfully interpreting the given data.
- Depending upon the purpose, models can be classified into four categories: Descriptive model, Explanatory model, Predictive model and Prescriptive (or normative) model.
- The main three types of averages commonly used are: Mean, Median, Mode.
- Quantitative techniques provide the executive with scientific basis for decision-making and enhance his ability to matte long-range plans and to solve every day problems of running a business and industry with greater efficiency and confidence.

1.9 Keywords

Descriptive Models: Models which are used to describe the behaviour of a system based on data.

Descriptive Statistics: It is concerned with the analysis and synthesis of data so that better description of the' situation can be made.

Explanatory Models: Models which are used to explain behaviour of a system by establishing relationships between its various components.

Inductive Statistics: It is concerned with the developments of scientific criteria which can be used to derive information about the group of data by examining only a small portion (sample) of that group.

Notes

Operations Research: It is a scientific method of providing executive departments with a quantitative basis for decision regarding the operations under control.

Predictive Models: Models which are used to predict the status of a system in the near future based on data.

Quantitative Techniques: It is the name given to the group of statistical and operations research (or programming) techniques.

Statistical Data: It refers to numerical description of quantitative aspects of things. These descriptions may take the form of counts or measurement.

Statistical Decision Theory: It is concerned with the establishment of rules and procedures for choosing the course of action from alternative courses of actions under situation of uncertainty.

Statistical Methods: These methods include all those devices of analysis and synthesis by means of which statistical data are systematically collected and used to explain or describe a given phenomenon.

Answers: Self Assessment

1.	Quantitative techniques	2.	Qualitative or judgment
3.	Statistical methods	4.	Secondary data
5.	Decision environment	6.	Linear Models
7.	Simulation model	8.	False
9.	True	10.	True
11.	True	12.	False

1.10 Review Questions

1. Briefly explain an overview of quantitative techniques. What factors in modern society contribute to the increasing importance of quantitative approach to management?
2. Describe the major phases of statistics. Formulate a business problem and analyse it by applying these phases. Also explain the various statistical methods.
3. Explain the distinction between:
   (a) Static and dynamic models
   (b) Analytical and simulation models
   (c) Descriptive and prescriptive models.
4. Describe the main features of the quantitative approach to management.
5. What is the need of using quantitative approach to managerial decisions?
6. What are the various models frequently used in operations research and the basis of their classification?
7. Explain the areas of applications of quantitative approach in business and management.

Notes

1.11 Further Readings

Books

Gupta, S.P. and M.P. Gupta, 1987. *Business Statistics*, Sultan Chand & Sons: New Delhi.

Loomba, M.P., 1978. *Management–A Quantitative Perspective*, MacMillan Publishing Company: New York.

Shenoy; G.V., U.K. Srivastava and S.C. Sharma, 1985. *Quantitative Techniques for Managerial Decision Making*, Wiley Eastern: New Delhi.

Venkata Rao, K., 1986. *Management Science*, McGraw-Hill Book Company: Singapore.

Online links

www.indiastudychannel.com

www.scribd.com/doc

www.soas.ac.uk

www.web-source.net

Unit 2: Introduction to Research

Notes

CONTENTS

Objectives

Introduction

2.1 Meaning of Research

2.1.1 Definition of Research

2.2 Features of a Good Research Study

2.3 Scope and Significance of Research

2.4 Goals, Strategy and Tactics of Research

2.5 Internal and External Research Suppliers

2.5.1 External Organizations for Conducting Marketing Research

2.6 Marketing Research – A Definition

2.6.1 Explanation

2.7 Scientific Method in Research

2.7.1 Characteristics of Scientific Method

2.7.2 Why MR cannot be considered Scientific

2.7.3 Distinction between Scientific and Unscientific Methods

2.7.4 Difficulties in Applying Scientific Methods to Marketing Research

2.8 Research Process

2.8.1 What is a Research Problem?

2.8.2 What is Research Methodology?

2.8.3 Research Design/Plan

2.8.4 Steps involved in Preparing Market Research Plan or Designing a Research

2.9 Formulating the Problem

2.9.1 Evaluate the Cost of Research

2.9.2 Preparing a List of Needed Information

2.9.3 Decision on Research Design

2.9.4 Select the Sample Types

2.9.5 Determine the Sample Size

2.9.6 Organize the Fieldwork

2.9.7 Analysis of the Data

2.10 Summary

2.11 Keywords

2.12 Review Questions

2.13 Further Readings

Notes

Objectives

After studying this unit, you will be able to:

- Know the meaning of research;
- Discuss the features of a good research study;
- Identify the goals, strategy and tactics of research;
- Explain the internal and external research suppliers.

Introduction

Research in management is particularly difficult because of its convergence with different disciplines. As we know, management is not a particular discipline and in any study on management we need to integrate the different approaches borrowing suitably from different disciplines. Similarly, before we understand the complexity of research in commerce and management, we need to define certain important concepts.

First is Variable. What exactly is a variable? Variable is the quantity, in which we are interested, that varies in the course of the research or that has different variables for different samples in our study. In one word, we can define variable as a factor whose change or difference we study. Now, there are two types of variables. The first one is Dependent variable and the second is Independent variable. Dependent variable is that quantity or aspect of nature whose change at different stages the researcher wants to understand or explain. In cause and effect investigation, the effect variable is the dependent variable.

Now, what exactly is an independent variable? Independent variable is a variable, whose effect upon the dependent variable we try to understand. There may be several independent variables. For instance, we may simultaneously investigate the effect of mother's cigarette smoking, mother's exercise, parents' weights and other variables upon the weight of a baby. In this case, mother's cigarette smoking, mother's exercise, parents' weights and other variables are independent variables, which we want to study, upon the weight of the baby, which is dependent variable.

Now, there are certain other areas. One such area is Universe. We can define it as the total population. It is the laboratory for the research. In our research we may have or we may take the entire population of India. In that case, as it is obvious, no researcher can carry out research on the entire population of India to find out the truth or to find out some areas of his research interest. In some cases, Universe or population may be a particular group. To clarify this point further, let us assume we want to study some effects on some particular group of people, religions; Hindu, Muslim, Christian, Jain, etc., or certain particular age-group, the age of 25 to 35. In that case our universe is getting limited to that particular religion or to that particular age-group of people. Similarly even then we find that carrying out research on the universe, i.e., on the entire population of that particular group may not be always possible because of the time factor and the money involved. In that case, what we usually do is to take out samples selected from the entire population. In selecting samples, we use the available sampling techniques to draw from the total population. Apart from these, we need to clarify certain other concepts.

Did u know? What does the term empirical mean?

Empirical means the observations and propositions which are primarily based on some sense experiments or derived from experience by methods of inductive logic including mathematics and statistics. This technical definition is difficult to understand. To be clearer we can define

Notes

empirical research as that type of research where we try to deduce some logic and principles based on our survey reports. In other words, when we want to analyze the survey report using some mathematical and statistical tools and deduce logic to authenticate our findings, we follow the empirical research method.

2.1 Meaning of Research

Research has been defined by various authors in different ways. It always begins with a question or a problem. Its purpose is to find answers to questions through the application of systematic and scientific methods. Thus, research is the systematic approach towards purposeful investigation. This needs formulating a hypothesis, collection of data on relevant variables, analyzing and interpreting the results and reaching conclusions either in the form of a solution or certain generalizations.

Notes Research is an academic activity and a systematized effort to gain new knowledge.

Research in common man's language refers to "search for knowledge".

Research is an art of scientific investigation. It is also a systematic design, collection, analysis and the reporting, the findings and solutions for the marketing problems of a company. Research is required because of the following reasons:

- To identify and find solutions to the problems
- To help making decisions
- To develop new concepts
- To find alternate strategies

To identify and find solutions to the problem

To understand the problem in depth. For example, "Why is that, demand for a product falling"? "Why business fluctuation takes place once in three years"? By identifying the problem precisely, it is easy to collect the relevant data to solve the problem.

- To help making decisions

Example: Should we maintain the same advertising budget as last year? Research will provide an answer to this question.

- To find alternative strategies

Example: Should we follow pull strategy or push strategy to promote the product?

- To develop new concepts

Example: CRM, Horizontal Marketing, MLM etc.

2.1.1 Definition of Research

Research may be defined as a documented prose work. Documented prose work means organized analysis of the subject based on borrowed materials with suitable acknowledgement and consultation in the main body of the paper. Research in management is particularly important to find out different phenomena.

Notes

Caution At the outset we should distinguish between researches in different areas.

Management research comes within the purview of social science research and there are other different types of research which broadly fall into the category of physical science research. Carrying out research in social science subjects, i.e., Commerce, Management, Economics, Sociology, etc., is basically different from Physical Science because, here we need to study the society based on certain trends and for this the laboratory is the society.

Self Assessment

Fill in the blanks:

1. ……is the systematic approach towards purposeful investigation.
2. Research is an art of ……investigation.

2.2 Features of a Good Research Study

Following are the features of a good research study:

i. ***Objectivity:*** A good research is objective in the sense that it must answer the research questions. This necessitates the formulation of a proper hypothesis; otherwise there may be lack of congruence between the research questions and the hypothesis.

ii. ***Control:*** A good research must be able to control all the variables. This requires randomization at all stages, e.g., in selecting the subjects, the sample size and the experimental treatments. This shall ensure an adequate control over the independent variables.

iii. ***Generalisability:*** We should be able to have almost the same result by using an identical methodology so that we can apply the result to similar situations.

iv. ***Free from Personal Biases:*** A good research should be free from the researcher's personal biases and must be based on objectivity and not subjectivity.

v. ***Systematic:*** A good research study must have various well planned steps, i.e., all steps must be interrelated and one step should lead to another step.

vi. ***Reproducible:*** A researcher should be able to get approximately the same results by using an identical methodology by conducting investigation on a population having characteristics identical to the one in the earlier study.

Hence, the following points must be ensured:

- Purpose clearly detailed
- Research design thoroughly planned
- High ethical standards applied
- Limitations frankly revealed
- A complete and proper analysis
- Findings presented unambiguously
- Decision based conclusions.

2.3 Scope and Significance of Research

Notes

Let us discuss the scope and significance of research:

i. ***Decision-making tool:*** Whenever a decision is to be made, business research becomes necessary in the corporate world. The degree of dependence on research is based on the cost of decisions. If the cost of decision is high, the dependence on research is high, and vice versa.

ii. ***Facilitates large-scale production:*** The MR helps large scale enterprises in the areas of production to determine:

   (a) What to produce?

   (b) How much to produce?

   (c) When to produce?

iii. ***To determine the pattern of consumption:*** The consumption patterns vary from place to place and time to time. The MR helps in identifying the consumption pattern and also the availability of consumer credit in that particular place.

   MR helps the marketer to identify:

   (a) Consumption pattern

   (b) Brand loyalty

   (c) Consumer behaviour

   (d) Market trends, etc.

iv. ***Complex market:*** In a complex and dynamic environment, the role of MR is very vital. MR acts as a bridge between the consumer and the purchaser. This is because MR enables the management to know the need of the customer, the about demand for the product and helps the producer to anticipate the changes in the market.

v. ***Problem-solving:*** The MR focuses on both short range and long range decisions and helps in making decisions with respect to the 4p's of marketing, namely, product, price, place and promotion.

vi. ***Distribution:*** The MR helps the manufacturer to decide about the channel, media, logistics planning so that its customers and distributors are benefited. Based on the study of MR, suitable distributors, retailers, wholesalers and agents are selected by the company for distributing their products.

vii. ***Sales promotion:*** The MR helps in effective sales promotion. It enlightens the manufacturer with regard to the method of sales promotion to be undertaken, such as advertising, personal selling, publicity etc. It also helps in understanding the attitude of the customers and helps how to design the advertisement in line with prevailing attitudes.

Self Assessment

Fill in the blanks:

3. A good research is ……..in the sense that it must answer the research questions.

4. …….acts as a bridge between the consumer and the purchaser.

Notes

2.4 Goals, Strategy and Tactics of Research

Many researchers agree that the goals of scientific research are: description, prediction, and explanation/understanding. Some individuals add control and application to the list of goals. The goal of research is to find out answers to questions through the application of systematic and scientific way.

Though there is a specific purpose behind each research study, however, the objectives can be broadly classified as under:

- To obtain familiarity of a phenomenon.
- To determine the association or independence of an activity.
- To determine the characteristics of an individual or a group of activities and the frequency of its (or their) occurrence.

2.5 Internal and External Research Suppliers

Marketing research can be conducted by having:

(a) Internal marketing departments in the organisation

or

(b) By taking the help of external agencies such as ORG, Marg, AC Neilson etc.

The type of organisation selected for market research depends on how big an organisation is and the varied type of products manufactured etc. There are two types of internal departments. Departments internal to organisation can be run by:

(a) One person

or

(b) A full-fledged market research department where several employees are involved.

One Person Operation

Small companies may not be able to afford a full-fledged MR department. They may appoint one or two persons to conduct MR and report the results to the head of the company. In medium sized firms, MR reports are collected by head of the marketing department. In larger firms, an independent marketing research department is established on a permanent basis and an experienced person is appointed as the head of marketing department.

The marketing department usually has executives, secretaries, assistants and others. The marketing department can function either on a centralized or decentralized basis. The centralized marketing department has the advantage of good coordination with various departments. On the other hand, decentralized marketing departments score in gaining valuable knowledge regarding markets, products in the respective area (i.e. is localised).

2.5.1 External Organizations for Conducting Marketing Research

(a) Advertising agencies

(b) Trade associations

(c) Manufacturers

Notes

(d) Retailers and wholesalers

(e) Governmental agencies

(f) Universities and institutions.

External Organisations – Advertising Agencies

The advertising agencies conduct marketing research for their clients. Ad agencies undertake media studies, group research etc. The ad agencies also conduct image opinion research, market potential research etc.

Example: The ad agencies that conduct MR are Hindustan Thompson, Mudra Communications, Maa Advertising et al.

Trade Associations

Trade associations also conduct MR. For instance, the Confederation of Engineering Industries conducts MR for various engineering products. In India, many consumer goods manufacturers such as Godrej, P&G, HLL have their own MR organizations.

Manufacturers

Manufacturers of smaller industries join together and undertake marketing research for their mutual benefit.

Example: Textile companies have the Textile Manufacturers Association, which conducts research on potential for garments made in the country.

Retailers and Wholesalers

The retailers have been predominantly concerned with shop location studies, special promotion studies, pricing, retail stores investigation and sales research and so on. The retailers also conduct a study on consumer behaviour and attitudes towards a particular product. The wholesalers are interested in conducting MR on retailers' behaviour. They are also interested in learning about the attitudes of the retailers towards inventories, service provided by the wholesalers etc. The latter category of researchers is lesser in number.

Government Agencies

MR is also carried out by few government agencies. The government departments collect information on subjects such as agricultural market surplus, consumer goods market surplus, price indices, imports and exports, etc. This helps them in formulating policies.

Universities and Institutions

Universities and institutions also conduct marketing research.

Example: The institutes like the IIMs, IIFT engage themselves in doing marketing research for certain corporate entities.

Notes

Task An Indian company dealing in pesticides hires a qualified business management graduate to expand its marketing activities. Most of the current employees of the company are qualified chemists with science background. During their first review meeting the management graduate says that the "company should be involved in market research to get a better perspective of the problem on hand". On hearing this, one of the science graduates laughs and says "There is no such thing as marketing or business research, research is combined to science alone." What would be your response?

Self Assessment

Fill in the blanks:

5. The type of organisation selected for …….depends on how big an organisation is and the varied type of products manufactured etc.
6. The goal of research is to find out answers to …….through the application of systematic and scientific way.

2.6 Marketing Research – A Definition

Marketing research involves;

(a) Systematic problem analysis

(b) Model building and

(c) Fact finding method, used for the purpose of important decision-making and to regulate the marketing of goods and services.

2.6.1 Explanation

From the above definition, it becomes clear that MR is the collection and interpretation of facts that help in marketing management to provide products and services more efficiently into the hands of consumers. It includes various types of research such as:

(a) Consumer research

(b) Sales research, etc.

2.7 Scientific Method in Research

Scientific research is one which yields the same results when repeated by different individuals. Scientific method consists of the following steps:

1. ***Observation:*** The researcher wants to observe a set of important factors that is related to his problem.
2. ***Formulates Hypothesis:*** The researcher formulates a hypothesis which will explain what he has observed.
3. ***Future Prediction:*** The researcher draws a logical conclusion.
4. ***Testing the Hypothesis:*** The researcher will arrive at the conclusion based on data.

Notes

Example: A simple example will highlight how a scientific method works.

Let us assume that a researcher is conducting a market research for a client manufacturing men's apparel.

1. *Observation:* The researcher observes that some of the competitors are doing brisk business. The increase on sales of apparel is mainly due to round or turtleneck shirt and narrow bottom pants.
2. *Formulation of Hypothesis:* The researcher now presumes that the products of his clients are somewhat similar and the variation in shirt and pant variety as above is the main cause for an increase in the sales of his competitors.
3. *Future Prediction:* It is predicted that if his client introduces similar products, the sales will increase.
4. *Hypothesis Testing:* The client now produces round-neck shirts and narrow bottom pants for test marketing.

2.7.1 Characteristics of Scientific Method

(a) Validity

(b) Reliability.

Validity is the ability of a measuring instrument to measure what it is supposed to. A questionnaire is administered to determine the attitudes of the respondent towards a movie. As long as the questionnaire serves this purpose, we say that the instrument is valid.

In physical sciences, the instruments used such as barometer, thermometer or foot ruler which measures what they are meant to do. Also, the measurement can be repeated any number of times by different individuals, but the result will be the same.

2.7.2 Why MR cannot be considered Scientific

In Marketing Research, the instrument used is a questionnaire. There are five main problems faced by researcher regarding validity and reliability:

1. Different respondents interpret the same question in different manner. So the reply of the respondents will be different.
2. It is difficult to ascertain whether the sample is a representative of the population or not.
3. The same questionnaire administered by different interviewers will yield different results.
4. The measuring instrument, namely the questionnaire may not state clearly what is being measured.
5. Lab experiments are held under controlled conditions, such as temperature, humidity etc. In marketing research, it is not possible to control external factors surrounding the study.

Example: The respondent is interviewed on a specific subject. After 60 days, the respondent is interviewed again reply could be very different from what he said earlier. This may happen because he gathered additional information, or had discussed the subject with others during this period.

Notes

Did u know? Reliability implies that we must obtain similar result again when measured.

Example: Linear measurements using a foot ruler, velocity of light and sound in a given media will be the same, when measured repeatedly.

2.7.3 Distinction between Scientific and Unscientific Methods

There are three differences between scientific and unscientific methods:

1. Rationality and objectivity
2. Accuracy of measurement
3. Maintaining continuity in investigation

Rationality and Objectivity

The conclusions should be based on facts. Our mindsets should not influence the decision-making.

Example: When the Howthorne studies began, it was thought that "employee satisfaction has improved productivity". Later research proved otherwise. In fact, subsequent research justified that productivity and employee satisfaction are not directly related.

Similarly, in marketing research, the researcher should not proceed with pre-conceived notions. He must keep an open mind and be objective. Sometimes, researchers approach the respondents, who are easy to reach, and with whom they are comfortable even though they may not represent the true sample. In this case, the objectivity is sacrificed.

Accuracy

Accuracy is possible through the use of scientific instruments. For, the measuring instrument is valid and reliable. In marketing research, a questionnaire is used to measure these aspects such as attitude, preference etc. but this instrument is crude.

Example: Habits such as smoking are measured using a scale like:

a. Often
b. Sometimes
c. More often than not
d. Rarely
e. Regularly

There are two aspects in the above questionnaire which may lead to inaccuracy:

1. 'Respondents' perception of what is asked
2. What is the correct answer among the alternative

It is difficult to judge whether the respondent is answering correctly. Due to all these factors, accuracy is often sacrificed.

Notes

Maintaining Continuity in Investigation

Science is marked by continuity. This is because, every time there is an invention, the same is carried forward for further improving the same.

Example: Basic telephony vs Latest mobile phones, early steam engines vs electrically driven engines.

In marketing research, there is less continuity. The present researcher does not start from where it was left off. Each project is independent. What is learnt in one assignment is not made use of in subsequent projects.

Due to all the above three reasons, we can conclude that marketing research is not scientific.

2.7.4 Difficulties in Applying Scientific Methods to Marketing Research

- Role of investigators
- Inaccuracy of measuring instruments
- Influence of measurement
- Pressures of time-frame
- Testing of hypothesis
- Complexity of the subject.

Role of Investigators

Organisations are the clients of researchers. Sometimes, the investigator tries to fit in results which are readily acceptable to clients. This is possible when the investigator manipulates the data or does not conduct an exhaustive study. In either of these circumstances, the study becomes unscientific.

Inaccuracy of Measuring Instruments

Accuracy of measurement separates scientific and unscientific methods. Since human beings are the participants, subjectivity invariably creeps in. Most of the information obtained from the respondent is qualitative in nature.

Example: Brand preference of a respondent. The respondent might say that "I immensely like this brand." It is difficult, if not impossible, to quantify this reply. The questionnaire used to measure attitudes is not precise.

Also, each of the interviewers will administer the questionnaire differently. Added to this, the respondent's casual answer may be due to pre-occupation, fatigue etc.

Influence of Measurement

In physical sciences, the researcher can repeat an experiment any time to get the same results. This is not the case with marketing research. When a respondent realises that he is being measured, his response and behaviour undergoes a change. Because, human reaction changes quickly. The reliability and validity of research will suffer a great deal.

Notes

Time Pressure

Marketing research must be conducted and completed within a given time-span. If more time is consumed in conducting the research, competitors might enter and capture the market. The research is concluded in a hurry, leading to lack of credibility due to the pressure exerted by the clients on the researcher.

Testing of Hypothesis

Any hypothesis formed in M.R must be tested. Thus, experimentation has to be resorted to. In marketing research, it is almost impractical to carryout experiment due to many factors which come in the way. For instance, while measuring the impact of advertising on sales, it may so happen that competitors have also advertise, resulting in lesser volume of sales. Also, it is impossible to reproduce the same experiment. The reliability of research suffers on account.

Complexity of the Subject

The subject becomes very complex, due to the fact that it is human beings, who interact with the researcher. Human reaction varies from time to time. Different individuals react differently for a given stimuli. Added to this are the environmental factors and peer influence adding to the complexity. Ads or promotions campaign held at different times yield different results due to changed perceptions and reactions by audience. Due to all these reasons, we can say that marketing research is complex.

Market researchers are subjected to pressure. The research is to be completed quickly before competitors enter the market. Due to this pressure, reliability might suffer and face difficulty in testing the hypothesis. Testing the hypothesis is the core in scientific research. In marketing, it is very difficult to control external factors in the field.

Example: Measuring the effect of advertising on consumer behaviour. Here, the advertising factor cannot be isolated from other factors such as change in the taste of customers or action taken by competitors.

Self Assessment

Fill in the blanks:

7. ……is the ability of a measuring instrument to measure what it is supposed to.
8. ……implies that we must obtain similar result again when measured.
9. ……....of measurement separates scientific and unscientific methods.

2.8 Research Process

Research process is the main content of marketing research. It defines marketing research and describes the skills required to identify the problem, the decision alternatives, and the client's needs, which are critical components of a research activities. The marketing research is expected to understand the market information needs of decision makers, and on the other hand expected to follow the proper processes and procedures for obtaining that information. Marketing research process involves a number of inter-related activities which overlap and do not rigidly follow a particular sequence. A researcher is often required to think a few steps ahead, because various steps in research process are inter-woven into each other and each step will have some influence

Notes

over the other steps. In marketing research, even though our focus is on one particular step, other inter-related steps of operations are also being looked into simultaneously. As we complete one activity or operation, our focus naturally shifts from it to the subsequent one, i.e. the focus is not concentrated exclusively on one single activity or operation at any particular point of time. The research process provides systematic, planned approach to the research project and ensures that all aspects of the research project are consistent with each other.

2.8.1 What is a Research Problem?

A research problem refers to some difficulty which an organisation faces and wishes to obtain a solution for the same.

Meaning and Definitions

Defining a research problem is the fuel that drives the scientific process, and is the foundation of any research method and experimental design, from true experiment to case study. The first and foremost step in the research process consists of problem or opportunity identification. The necessity of properly identified research problems cannot be overemphasized. It is rightly said that a problem properly defined is half solved.

Based upon the objective, the research problem could be in any of the following three areas:

i. Exploratory for gathering preliminary information that may help in defining the problem and suggest hypothesis. The major emphasis of exploratory research is on the discovery of ideas. The idea is to clarify concepts and subsequently make more extensive research on them.

ii. Descriptive, which may describe things such as market potential for a product or the demographics and attitudes of a customer who buys the product.

iii. Casual, to test hypothesis about cause and effect relationships.

Once the researcher has identified two or more problems or opportunities, the next question for him is to select a problem based on priority, limited finance and time constraints. He should choose the problem which is likely to add value to the research. Choosing a relatively less important problem would amount to wasting time and resources.

Initially, the problem may be stated in a broad general way and then the clarifications if any, can be resolved as the research advances. The researcher must, at the same time, examine all available literature to get himself acquainted with the selected problem.

While doing research, defining the problem is very important because "problem clearly stated is half-solved". This shows how important it is to "define the problem correctly". While defining the problem, it should be noted that definition should be unambiguous. If the problem defining is ambiguous, then the researcher will not know "what data is to be collected" or "what technique is to be used" etc.

Example: An ambiguous definition: "Find out by how much sales have declined recently".

Let us suppose that the research problem is defined in a broad and general way as follows:

"Why is the productivity in Korea much higher than that in India"? In this type of question, a number of ambiguities are there, such as:

- What sort of productivity is to be specified; is it men, machine, materials?
- To which type of industry is the productivity related to?
- In which time-frame are we analyzing the productivity?

Notes

Example: An unambiguous definition: On the contrary, a problem will be as follows:

"What are the factors responsible for increased labour productivity in Korean textile manufacturing industries during 1996-07 relative to Indian textile industries?"

2.8.2 What is Research Methodology?

Research methodology is a method to solve the research problem systematically. It involves gathering data, use of statistical techniques, interpretations, and drawing conclusions about the research data. It is a blueprint, which is followed to complete the study. It is similar to builders' blue-print for building a house.

2.8.3 Research Design/Plan

Research design is one of the important steps in marketing research. It helps in establishing the manner researchers go about to achieve the objective of the study. The preparation of a research design involves a careful consideration of the following questions and making appropriate decisions about them:

1. What the study is about?
2. Why is the study undertaken?
3. What is its scope?
4. What are the objectives of the study?
5. What are the hypotheses/proportions to be tested?
6. What are the major concepts to be defined operationally?
7. What type of literature needs to be reviewed?
8. What is the area of study?
9. What is the reference period of study?
10. What is the methodology to be used?
11. What kinds of data are needed?
12. What are the sources of data?
13. What is the sampling boundary?
14. What are the sampling units?
15. What is the sample size?
16. What are the sampling techniques?
17. What are the data collection methods?
18. How is the data processed?
19. What are the statistical techniques for analysis?
20. What is the target group, the finding are meant for?
21. What is the type of report?
22. What is the duration of time required for each stage of the research work?

Notes

23. What is the cost involved?

24. Who reads the report?

2.8.4 Steps involved in Preparing Market Research Plan or Designing a Research

There are nine steps in the research process that can be followed while designing a research project. They are as follows:

- Formulate the problem
- Evaluate the cost of research
- Prepare the list of information
- Research design decision
- Data collection
- Select the sample type
- Determine the sample size
- Organize the field work
- Analyze the data and report preparation.

Self Assessment

Fill in the blanks:

10. The …….provides systematic, planned approach to the research project and ensures that all aspects of the research project are consistent with each other.

11. Marketing research process involves a number of ……activities which overlap and do not rigidly follow a particular sequence.

2.9 Formulating the Problem

Problem formulation is the key to research process. For a researcher, the problem formulation means converting the management problem to a research problem. In order to attain clarity, the MR Manager and the researcher must articulate clearly so that perfect understanding of each other is achieved.

Example: Management problem and research problem

M.P - Want to increase the sale of product A

R.P - What is the current standing of the product A?

While problem is being formulated, the following should be taken into account:

(1) Determine the objective of the study.

(2) Consider various environmental factors.

(3) Nature of the problem.

(4) Stating the alternative.

Notes

1. ***Determine the objective:*** The objective may be general or specific. General category – It would like to know how effective the advertising campaign was.

   The corollary looks like a statement with an objective. In reality, this is far from the case. There are two ways of determining the objectives precisely: (1) The researcher should clarify with the MR manager "what effective means". Does effective mean, the awareness or does it refer to an increase in sales or does it mean it has improved the knowledge of the audience, or the perception of audience about the product? In each of the above circumstances, the question to be asked from the audience varies (2) Another way to determine objectives is to find out from the MR Manager, "What action will be taken, given the specified outcome of the study?"

   Example: If research findings to the previous advertisement by the company was indeed ineffective, what course of action does the company intend to take? (a) Increase the budget for the next Ad (b) Use different appeal (c) Change the media (d) Go to a new agency.

   If the objectives are proper, the research questions will be precise. However, we should remember that objectives do undergo a change.

2. ***Consider environmental factors:*** Environmental factors influence the outcome of the research and the decision. Therefore, the researcher must help his client to identify the environmental factors that are relevant.

   Example: Assume that the company wants to introduce a new product like iced tea or frozen green peas or ready to eat chapathis.

   The following environmental factors are to be considered:

   1. Purchasing habits of consumers
   2. Presently, who are the competitors in the market with similar product.
   3. What is the perception of the people about other products of the company, with respect to price, image of the company.
   4. Size of the market and target audience.

   All the above factors could influence the decision. Therefore, the researcher must work very closely with his client.

3. ***Nature of the problem:*** By understanding the nature of the problem, the researcher can collect relevant data and help suggest a suitable solution. Every problem is related to either one or more variables. Before beginning the data collection, a preliminary investigation of the problem is necessary for a better understanding of the same.

> *Notes* Initial investigation could be carried by using a focus group of consumers or sales representatives.

   If a focus group is carried out with consumers, some of the following questions will help the researcher to understand the problem better:

   i. Did the customer ever include this company's product in his mental map?

   ii. If the customer is not buying the company's product, the reasons for his not doing so.

Notes

iii. Why did the customer turn to the competitor's product?

iv. Is the researcher contacting the right target audience?

4. ***Stating the alternatives:*** The researcher would be better served by generating as many alternatives as possible during the problem formulation hypothesis.

Example: Whether to introduce a sachet form of packaging with a view to increase sales. The hypothesis may state that acceptance of the sachet by the customer will increase the sales by 20%. Thereafter, the test marketing will be conducted before deciding whether to introduce the sachet variant. Therefore, for every alternative, a hypothesis has to be developed.

2.9.1 Evaluate the Cost of Research

There are several methods to establish the value of research. Some of them are (1) Bayesian approach (2) Simple saving method (3) Return on investment (4) Cost benefit approach.

Example: Company 'X' wants to launch a product. The company's intuitive feeling is that the possibilities of the product's failure are 35%. However, if research is conducted and appropriate data is gathered, the chances of failure could be reduced to 30%. The company has calculated that losses would be to the tune of ₹ 3,00,000 if the product fails. The company has received a quotation from an MR agency. The cost of the intended research is ₹ 75,000. The question is: "Should the company spend this money to conduct the research?"

Calculation:

Loss without research = ₹ 3,00,000 × 0.35

= ₹ 1,05,000

Loss with research = ₹ 3,00,000 × 0.30

= ₹ 90,000

Value of research information

= ₹ 1,05,000 – 90,000

= ₹ 15,000

Since the value of information, namely ₹ 15,000 is lower than the cost of research, i.e., ₹ 75,000, conducting this particular research is not recommended.

Example: Company 'A' would like to introduce a new product in the market. The research agencies have given an estimation of ₹ 5 lakhs and a time period of five months. According to the past experience of the company, the probability of earning ₹ 10 lakh is 0.4 and ₹ 5 lakh is 0.3 and losing ₹ 7 lakh is 0.3. Should the company undertake the research?

Calculation:

0.4 × ₹ 10 + 0.3 × ₹ 5 – 0.3 × ₹ 7 = ₹ 4 + ₹ 1.5 – ₹ 2.1 = 3.4 lakh

Since we find that the expected value of information i.e. ₹ 3.4 lakh, less than the cost of M.R at ₹ 5 lakh, there is no need to carry out this research.

Notes

2.9.2 Preparing a List of Needed Information

Assume that company 'X' wants to introduce a new product (tea powder). Before introducing this product, it has to be test marketed. The company needs to know the extent of competition, price and quality acceptance from the market. In this context, following is the list of information required:

(a) ***Total demand and company sales***

Example: What is the overall industry demand? What is the share of competitors? The above information will help the management estimate the overall share and its own share in the market.

(b) ***Distribution coverage***

Example:

(1) Availability of products at different outlets.

(2) Effect of shelf display on sales.

(c) ***Market awareness, attitudes and usage***

Example:

"What percentage of target population is aware of the firm's product"?

"Do customers know about the product"?

"What is the customers' attitude towards the product"?

"What percentages of customers repurchase the product"?

(d) ***Marketing expenditure***

Example: "What has been the marketing expenditure"?

"How much was spent on promotion"?

(e) ***Competitors' marketing expenditure***

Example: "How much did the competitor spend to market a similar product"?

2.9.3 Decision on Research Design

Should the research be exploratory or conclusive?

Exploratory research:

Example: "Causes for the decline in sales of a specific company's product in a specific territory under a specific salesman".

The researcher may explore possible reasons as to why sales are failing.

- Faulty product planning
- Higher price

Notes

- Less discount
- Less availability
- Inefficient advertising/salesmanship
- Poor quality of salesman ship
- Less awareness

Not all factors are responsible for decline in sales.

Conclusive research: Narrow down the option. Only one or two factors are responsible for decline in sales. Therefore zero down, and use judgment and past experience.

(a) *Who should be interviewed for collecting data?:* If the study is undertaken to determine whether children influence the brand, for ready - to eat cereal (corn flakes) purchased by parents. The researcher must decide, if only adults are to be studied or children too included. The researcher must decide if data is to be collected by observation method or by interviewing. If an interview is chosen, should it be a personal interview or telephonic interview or questionnaire?

(b) *Should a few cases be studied or a large sample be chosen?:* The researcher may feel that there are some cases available which are identical and similar in nature. He may decide to use these cases for formulating the initial hypothesis. If suitable cases are not available, then the researcher may decide to choose a larger sample.

(c) *How to incorporate experiment in research?:* In an experiment, it has to be decided at the outset as to where and when measurements are to be conducted.

Example: In a test of advertising copy, the respondents can first be interviewed to measure their present awareness, and their attitudes towards certain brands. Then, they can be shown a pilot version of the proposed advertisement copy. Following this, their attitude too has to be measured again, to see if the proposed copy had any effect on them.

If it is a questionnaire, then the following questions should be postal– (a) What are the contents of the questionnaire? (b) What type of questions are to be asked? Pointed questions, general questions etc. (c) In what sequence should the questions be asked? (d) Should there be a fixed set of alternatives or should the question be open-ended. (e) Should the purpose be made clear to the respondents or should the same be disguised? are to be determined well in advance.

2.9.4 Select the Sample Types

The first task is to carefully select which groups of people or stores are to be sampled.

Example: Collecting the data from a fast food chain. Here, it is necessary to define what is meant by fast food chain. Also, the precise geographical location should be mentioned.

The next step is to decide whether to choose probability sampling or non-probability sampling. Probability sampling is one in which each element has a known chance of being selected. A non-probability sampling can be convenience or judgment sampling.

Notes

2.9.5 Determine the Sample Size

Smaller the sample size, larger the error and vice-versa.

Sample size depends upon

(a) Accuracy required

(b) Time available

(c) Cost involved.

Caution While selecting the sample, the sample unit has to be clearly specified.

Example: Survey on the attitudes towards the use of shampoo with reference to a specific brand, where husbands, wives or a combination of them are to be surveyed or a specific segment is to be surveyed. The sample size depends on the size of the sample frame/universe.

2.9.6 Organize the Fieldwork

This includes selection, training and evaluating the field sales force to collect the data:

(a) How to organise the field-work?

(b) What type of questionnaire - structured or unstructured to use?

(c) How to approach the respondents?

(d) Week, day and time to meet the specific respondents etc., are to be decided.

2.9.7 Analysis of the Data

This involves:

(a) Editing

(b) Tabulating

(c) Codifying.

Editing: The data collected should be scanned to make sure that it is complete and that all the instructions are followed. This process is called editing. Once these forms have been edited, they must be coded.

Coding means assigning numbers to each of the answers, so that they can be analysed.

The final step is called data tabulation. It is the orderly arrangement data in a tabular form. Also, at the time of analysing the data, the statistical tests to be used must be finalised such as T-Test, Z-Test, Chi-square Test, ANOVA etc.

Tasks

A. Given the following decision problem, identify the research problem:

1. Whether to expand the available warehouse facilities.

Contd...

Notes

2. Whether to change the compensation package of the sales force.
3. Whether to increase the expenditure on print advertisement.

B. Given the following research problem, identify the corresponding decision problem for which the information will be useful:

1. Assess the level of awareness among housewives regarding the benefits of introducing a new product in the market.
2. Assess attitudes and opinions of customers towards existing five-star hotels.
3. Design a test market to assess the effect of particular discount scheme on the volume of sales of the product.

Self Assessment

Fill in the blanks:

12. While selecting the sample, the ……..has to be clearly specified.
13. ……..means assigning numbers to each of the answers, so that they can be analysed.

2.10 Summary

- Research originates in a decision process.
- In research process, management problem is converted into a research problem which is the major objective of the study.
- The goal of research is to find out answers to questions through the application of systematic and scientific way.
- Scientific research is one which yields the same results when repeated by different individuals.
- Research process is the main content of marketing research.
- A research problem refers to some difficulty which an organisation faces and wishes to obtain a solution for the same.
- Defining a research problem is the fuel that drives the scientific process, and is the foundation of any research method and experimental design, from true experiment to case study.
- Research methodology is a method to solve the research problem systematically.
- Problem formulation is the key to research process. For a researcher, the problem formulation means converting the management problem to a research problem.

2.11 Keywords

Research Problem: A research problem refers to some difficulty which an organisation faces and wishes to obtain a solution for the same.

Research Process: Research process is the main content of marketing research. It defines marketing research and describes the skills required to identify the problem, the decision alternatives, and the client's needs, which are critical components of a research activities.

Research: Research is an art of scientific investigation. It is also a systematic design, collection, analysis and the reporting the findings and solutions for the marketing problems of a company.

Notes

Validity: Validity is the ability of a measuring instrument to measure what it is supposed to.

Variable: Variable is the quantity, in which we are interested, that varies in the course of the research or that has different variables for different samples in our study.

Answers: Self Assessment

1.	Research	2.	Scientific
3.	Objective	4.	MR
5.	Market research	6.	Questions
7.	Validity	8.	Reliability
9.	Accuracy	10.	research process
11.	inter-related	12.	Sample unit
13.	Coding		

2.12 Review Questions

1. What do you mean by research?
2. What is the importance of research?
3. Write down the features of a good research study.
4. Explain the goals, strategy and tactics of research.
5. Who are the internal and external research suppliers?

2.13 Further Readings

Books

Arthur, Maurice, *Philosophy of Scientific Investigation*, Baltimore: John Hopkins University Press, 1943.

Bernal, J.D., *The Social Function of Science*, London: George Routledge and Sons, 1939.

Chase, Stuart, *The Proper Study of Mankind: An Inquiry into the Science of Human Relations*, New York, Harper and Row Publishers, 1958.

S. N. Murthy and U. Bhojanna, *Business Research Methods*, Excel Books, 2007.

Online links

www.indiastudychannel.com

www.scribd.com/doc

www.soas.ac.uk

www.web-source.net

Notes

Unit 3: Language of Research

CONTENTS

Objectives
Introduction
3.1 Construct
3.2 Definitions/Concept
3.2.1 Types of Questions
3.2.2 Time in Research
3.2.3 Types of Relationships
3.3 Variables
3.3.1 Meaning
3.3.2 Attribute
3.3.3 Dependency
3.3.4 Exhaustive
3.4 Propositions-Hypotheses Research Process
3.4.1 Hypothesis Testing
3.5 Summary
3.6 Keywords
3.7 Review Questions
3.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss the construct of research;
- Know the concept of research;
- Identify the variables;
- Describe the propositions and hypotheses research process.

Introduction

Learning about research is a lot like learning about anything else. To start, we need to learn the jargon people use, the controversies they fight over, and the different factions that define the major players.

If the language is not stuck to, it might just be a surprise to see how esoteric the discussion can get (but not enough to cause you to give up in total despair). The language of research also includes some of the major issues in research like the types of questions we can ask in a project, the role of time in research, and the different types of relationships we can estimate. Then we have to consider defining some basic terms like variable, hypothesis, data, and unit of analysis.

Notes

Research involves an eclectic blending of an enormous range of skills and activities. To be a good social researcher, you have to be able to work well with a wide variety of people, understand the specific methods used to conduct research, understand the subject that you are studying, be able to convince someone to give you the funds to study it, stay on track and on schedule, speak and write persuasively, and on and on.

3.1 Construct

1. ***Theoretical and Empirical:*** A research might be theoretical or empirical depending on the way it is written and the objective it needs to follow.

   If much of a research is concerned with developing, exploring or testing the theories or ideas that social researchers have about how the world operates, then it is known as theoretical research. But sometimes, the research can also be empirical, meaning that it is based on observations and measurements of reality – on what we perceive of the world around us. You can even think of most research as a blending of these two terms – a comparison of our theories about how the world operates with our observations of its operation

2. ***Nomothetic and Idiographic:*** The word nomothetic comes perhaps from the writings of the psychologist Gordon Allport. It refers to laws or rules that pertain to the general case (nomos in Greek) and is contrasted with the term "idiographic" which refers to laws or rules that relate to individuals (idios means 'self' or 'characteristic of an individual in Greek).

Notes In any event, the point is that most social research is concerned with the nomothetic– the general case – rather than the individual. We often study individuals, but usually we are interested in generalizing to more than just the individual.

3. ***Probabilistic and Realistic:*** In our post-positivist view of science, we no longer regard certainty as attainable. Thus probabilistic as a term represents much contemporary social research which is most often than not, based on probabilities. The inferences that we make in social research have probabilities associated with them – they are seldom meant to be considered covering laws that pertain to all cases. Part of the reason we have seen statistics become so dominant in social research is that it allows us to estimate probabilities for the situations we study.

3.2 Definitions/Concept

Every research involves certain mandatory concepts that have to be understood and well applied by every researcher.

3.2.1 Types of Questions

There are three basic types of questions that research projects can address:

1. ***Descriptive:*** When a study is designed primarily to describe what is going on or what exists. Public opinion polls that seek only to describe the proportion of people who hold various opinions are primarily descriptive in nature.

Example: If we want to know what percent of the population would vote for a Democratic or a Republican in the next presidential election, we are simply interested in describing something.

Notes

2. ***Relational:*** When a study is designed to look at the relationships between two or more variables.

Example: A public opinion poll that compares what proportion of males and females say they would vote for a Democratic or a Republican candidate in the next presidential election is essentially studying the relationship between gender and voting preference.

3. ***Causal:*** When a study is designed to determine whether one or more variables (e.g., a program or treatment variable) causes or affects one or more outcome variables.

Example: If we did a public opinion poll to try to determine whether a recent political advertising campaign changed voter preferences, we would essentially be studying whether the campaign (cause) changed the proportion of voters who would vote Democratic or Republican (effect).

The three question types can be viewed as cumulative. That is, a relational study assumes that you can first describe (by measuring or observing) each of the variables you are trying to relate. And, a causal study assumes that you can describe both the cause and effect variables and that you can show that they are related to each other. Causal studies are probably the most demanding of the three.

3.2.2 Time in Research

Time is an important element of any research design; let us discuss one of the most fundamental distinctions in research design nomenclature:

Cross-sectional versus Longitudinal Studies

A cross-sectional study is one that takes place at a single point in time. In effect, we are taking a 'slice' or cross-section of whatever it is we're observing or measuring. A longitudinal study is one that takes place over time – we have at least two (and often more) waves of measurement in a longitudinal design.

A further distinction is made between two types of longitudinal designs: Repeated measures and time series.

There is no universally agreed upon rule for distinguishing these two terms, but in general, if you have two or a few waves of measurement, you are using a repeated measures design. If you have many waves of measurement over time, you have a time series. How many is 'many'? Usually, we wouldn't use the term time series unless we had at least twenty waves of measurement, and often far more. Sometimes the way we distinguish these is with the analysis methods we would use. Time series analysis requires that you have at least twenty or so observations. Repeated measures analyses (like repeated measures ANOVA) aren't often used with as many as twenty waves of measurement.

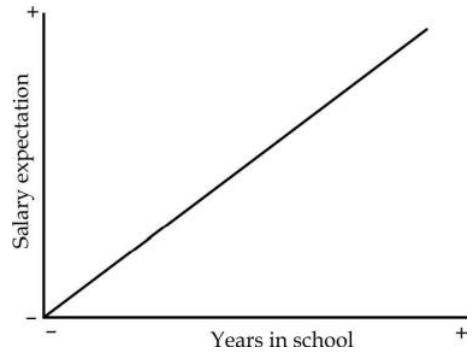
3.2.3 Types of Relationships

A relationship refers to the correspondence between two variables. When we talk about types of relationships, we can mean that in at least two ways: the nature of the relationship or the pattern of it.

Notes

The Nature of a Relationship

Figure 3.1



While all relationships tell about the correspondence between two variables, there is a special type of relationship that holds that the two variables are not only in correspondence, but that one causes the other. This is the key distinction between a simple correlational relationship and a causal relationship. A correlational relationship simply says that two things perform in a synchronized manner.

For instance, we often talk of a correlation between inflation and unemployment. When inflation is high, unemployment also tends to be high. When inflation is low, unemployment also tends to be low. The two variables are correlated. But knowing that two variables are correlated does not tell us whether one causes the other. We know, for instance, that there is a correlation between the number of roads built in Europe and the number of children born in India. Does that mean that is we want fewer children in India, we should stop building so many roads in Europe? Or, does it mean that if we don't have enough roads in Europe, we should encourage Indian citizens to have more babies? Of course not. While there is a relationship between the number of roads built and the number of babies, we don't believe that the relationship is a causal one. This leads to consideration of what is often termed the third variable problem. In this example, it may be that there is a third variable that is causing both the building of roads and the birthrate that is causing the correlation we observe. For instance, perhaps the general world economy is responsible for both. When the economy is good more roads are built in Europe and more children are born in India. The key lesson here is that you have to be careful when you interpret correlations. If you observe a correlation between the number of hours students use the computer to study and their grade point averages (with high computer users getting higher grades), you cannot assume that the relationship is causal: that computer use improves grades. In this case, the third variable might be socioeconomic status – richer students who have greater resources at their disposal tend to both use computers and do better in their grades. It's the resources that drive both use and grades, not computer use that causes the change in the grade point average.

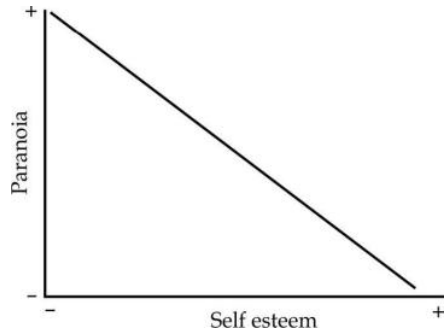
Patterns of Relationships

We have several terms to describe the major different types of patterns one might find in a relationship. First, there is the case of no relationship at all. If you know the values on one variable, you don't know anything about the values on the other.

Then, we have the positive relationship. In a positive relationship, high values on one variable are associated with high values on the other and low values on one are associated with low values on the other. In this example, we assume an idealized positive relationship between years of education and the salary one might expect to be making.

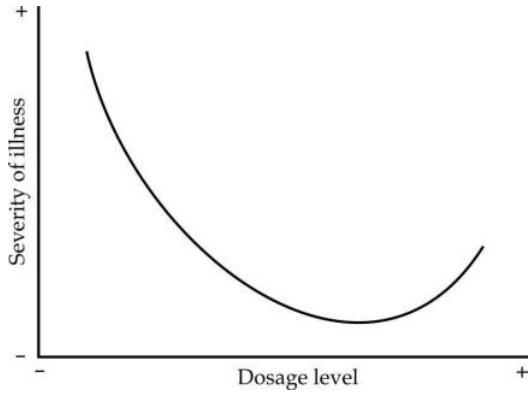
Notes

Figure 3.2



On the other hand a negative relationship implies that high values on one variable are associated with low values on the other. This is also sometimes termed an inverse relationship. Here, we show an idealized negative relationship between a measure of self esteem and a measure of paranoia in psychiatric patients.

Figure 3.3



These are the simplest types of relationships we might typically estimate in research. But the pattern of a relationship can be more complex than this. For instance, the figure on the left shows a relationship that changes over the range of both variables, a curvilinear relationship. In this example, the horizontal axis represents dosage of a drug for an illness and the vertical axis represents a severity of illness measure. As dosage rises, severity of illness goes down. But at some point, the patient begins to experience negative side effects associated with too high a dosage, and the severity of illness begins to increase again.

Self Assessment

Fill in the blanks:

1. Probabilistic as a term represents much contemporary social research which is most often than not, based on ……..
2. A ……..is one that takes place over time – we have at least two (and often more) waves of measurement in a longitudinal design.
3. A …….refers to the correspondence between two variables.
4. A ……relationship implies that high values on one variable are associated with low values on the other.

Notes

3.3 Variables

You won't be able to do very much in research unless you know how to talk about variables.

3.3.1 Meaning

A variable is any entity that can take on different values. OK, so what does that mean? Anything that can vary can be considered a variable.

Example: Age can be considered a variable because age can take different values for different people or for the same person at different times.

Similarly, country can be considered a variable because a person's country can be assigned a value.

Caution Variables aren't always 'quantitative' or numerical. The variable 'gender' consists of two text values: 'male' and 'female'. We can, if it is useful, assign quantitative values instead of (or in place of) the text values, but we don't have to assign numbers in order for something to be a variable.

It's also important to realize that variables aren't only things that we measure in the traditional sense. For instance, in much social research and in program evaluation, we consider the treatment or program to be made up of one or more variables (i.e., the 'cause' can be considered a variable). An educational program can have varying amounts of 'time on task', 'classroom settings', 'student-teacher ratios', and so on. So even the program can be considered a variable (which can be made up of a number of sub-variables).

3.3.2 Attribute

An attribute is a specific value on a variable. For instance, the variable sex or gender has two attributes: male and female. Or, the variable agreement might be defined as having five attributes:

1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree

Did u know? **What is the difference between attributes and variables?**

In contrast to variables, which are intended for bulk data, attributes are intended for ancillary data, or information about the data. The total amount of ancillary data associated with a net CDF object, and stored in its attributes, is typically small enough to be memory-resident. However variables are often too large to entirely fit in memory and must be split into sections for processing.

Another difference between attributes and variables is that variables may be multidimensional. Attributes are all either scalars (single-valued) or vectors (a single, fixed dimension).

Notes

Variables are created with a name, type, and shape before they are assigned data values, so a variable may exist with no values. The value of an attribute is specified when it is created, unless it is a zero-length attribute.

A variable may have attributes, but an attribute cannot have attributes. Attributes assigned to variables may have the same units as the variable (for example, valid-range) or have no units (for example, scale-factor). If you want to store data that requires units different from those of the associated variable, it is better to use a variable than an attribute.

3.3.3 Dependency

Another important distinction having to do with the term 'variable' is the distinction between an independent and dependent variable. This distinction is particularly relevant when you are investigating cause-effect relationships.

The terms dependent and independent variables are used to distinguish between two types of quantities being considered, separating them into those available at the start of a process and those being created by it, where the latter (dependent variables) are dependent on the former (independent variables).

In a research experiment, the dependent variable (DV) is the event studied and expected to change whenever the independent variable is altered.

In the design of experiments, an independent variable's values are controlled or selected by the experimenter to determine its relationship to an observed phenomenon (i.e., the dependent variable). In such an experiment, an attempt is made to find evidence that the values of the independent variable determine the values of the dependent variable. The independent variable (IV) can be changed as required, and its values do not represent a problem requiring explanation in an analysis, but are taken simply as given.

Caution The dependent variable, usually cannot be directly controlled.

Controlled variables are also important to identify in experiments. They are the variables that are kept constant to prevent their influence on the effect of the independent variable on the dependent. Every experiment has a controlling variable, and it is necessary to not change it, or the results of the experiment won't be valid.

"Extraneous variables" are those that might affect the relationship between the independent and dependent variables. Extraneous variables are usually not theoretically interesting. They are measured in order for the experimenter to compensate for them.

Example: An experimenter who wishes to measure the degree to which caffeine intake (the independent variable) influences explicit recall for a word list (the dependent variable) might also measure the participant's age (extraneous variable). He can then use these age data to control for the uninteresting effect of age, clarifying the relationship between caffeine and memory.

In summary:

1. Independent variables answer the question "What do I change?"
2. Dependent variables answer the question "What do I observe?"
3. Controlled variables answer the question "What do I keep the same?"

Notes

4. Extraneous variables answer the question "What uninteresting variables might mediate the effect of the IV on the DV?"

3.3.4 Exhaustive

Finally, there are two traits of variables that should always be achieved. Each variable should be exhaustive, it should include all possible answerable responses.

Example: If the variable is "religion" and the only options are "Christians", Hindus", "Jewish", and "Muslim", there are quite a few religions that haven't been included. The list does not exhaust all possibilities. On the other hand, if you exhaust all the possibilities with some variables – religion being one of them – you would simply have too many responses.

The way to deal with this is to explicitly list the most common attributes and then use a general category like "Other" to account for all remaining ones. In addition to being exhaustive, the attributes of a variable should be mutually exclusive; no respondent should be able to have two attributes simultaneously. While this might seem obvious, it is often rather tricky in practice.

Example: You might be tempted to represent the variable "Employment Status" with the two attributes "employed" and "unemployed." But these attributes are not necessarily mutually exclusive – a person who is looking for a second job while employed would be able to check both attributes! But don't we often use questions on surveys that ask the respondent to "check all that apply" and then list a series of categories? Yes, we do, but technically speaking, each of the categories in a question like that is its own variable and is treated dichotomously as either "checked" or "unchecked", attributes that are mutually exclusive.

Task Give three examples of events in which the variables are mutually exclusive.

Self Assessment

Fill in the blanks:

5. A ……..is any entity that can take on different values.
6. An ……is a specific value on a variable.
7. ……variable are those that might affect the relationship between the independent and dependent variables.

3.4 Propositions-Hypotheses Research Process

A hypothesis is a proposition – a tentative assumption which a researcher wants to test for its logical or empirical consequences. Hypotheses are more useful when stated in precise and clearly defined terms. It may be mentioned that though a hypothesis is useful it is not always necessary, especially in case of exploratory researches. However, in a problem-oriented research, it is necessary to formulate a hypothesis or hypotheses. In such researches, hypotheses are generally concerned with the causes of a certain phenomenon or a relationship between two or more variables under investigation.

Notes

3.4.1 Hypothesis Testing

A number of steps are involved in testing a hypothesis:

i. Formulate a hypothesis

ii. Set up a suitable significance level

iii. Choose a test criterion

iv. Compute the statistic

v. Make decision.

- ***Formulate a Hypothesis:*** Let us discuss about introduction of a new drug. The drug is tested on a few patients and based on the response from patients, a decision has to be made whether the drug should be introduced or not. We make certain assumptions about the parameter to be tested – these assumptions are known as hypotheses.

  We start with a 'null hypothesis': H_0 : m = 100. This is a claim or hypothesis about the values or population parameter.

  This is tested against alternate hypothesis, H_1 : m $\neq$ 100.

  The null hypothesis is tested with available evidence and a decision is made whether to accept this hypothesis or reject it. If the null hypothesis is rejected, we accept the alternate hypothesis.

- ***Setting up a Suitable Significance Level:*** There are two types of errors that can be committed in making decisions regarding accepting or rejecting the null hypothesis:

  - *Type I error:* An error made in rejecting the null hypothesis, when in fact it is true.
  - *Type II error:* An error made in accepting the null hypothesis, when in fact it is untrue.

Did u know? **What is level of significance?**

The level of significance signifies the probability of committing Type I error and is generally taken as equal to 5% (a = .05).

This means that even after testing the hypothesis, when a decision is made, we may still be committing 5% error in rejecting the null hypothesis when it is actually true. Sometimes, the value of 'a' is taken as .01 but it is the discretion of the investigator, depending upon the sensitivity of the study.

- ***Choose a Test Criterion:*** This means selection of a suitable test statistic that can be used along with the available information carrying out the test. The different test statistics that are normally used are:

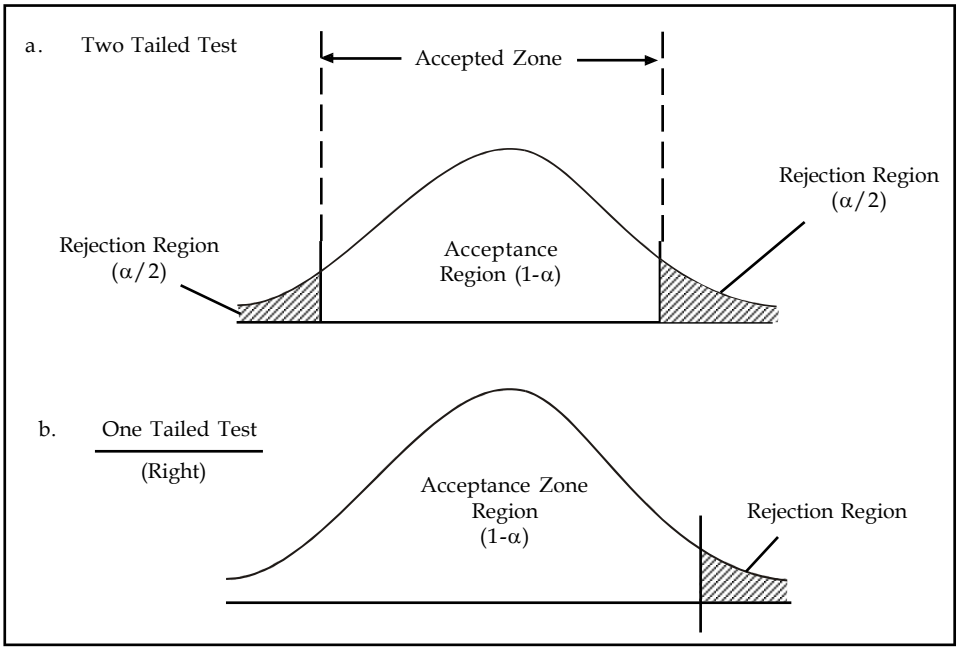
  - Normal Distribution: z-statistic, this is most often used, when the samples are more than 30.
  - t-statistic: 't' test is used for small samples only.
  - F-statistic
  - Chi-Square statistic.

- ***Compute the Test Characteristic:*** This involves the actual collection and computation of the sample data. For the case under consideration, we have to find the sample mean () and

Notes

then compute the calculated 'Z'. This calculated value (absolute) is compared with tabulated value obtained from normal distribution table against the decided criterion (value of 'a' and one tail or two tails).

- ***Make a Decision:*** If the calculated value of the test characteristic is greater than the tabulated value, the null hypothesis is rejected and the alternate hypothesis is accepted. Talking in terms of critical region, the value of calculated characteristic falls outside the acceptance region.



Task In the given following research problem, identify the corresponding decision problem for which the information will be useful:

1. Assess the level of awareness among housewives regarding the benefits of a new product to be introduced in the market.
2. Assess attitudes and opinions of customers towards existing five-star hotels.
3. Design a test market to assess the effect of particular discount scheme on the volume of sales of the product.

Self Assessment

Fill in the blanks:

8. The …….is tested with available evidence and a decision is made whether to accept this hypothesis or reject it.
9. ……means selection of a suitable test statistic that can be used along with the available information carrying out the test.

Notes

3.5 Summary

- The first thing that a good researcher needs to have is the language of research.
- If one doesn't, it is for sure that one is going to have a hard time discussing research.
- One has to take care of some of the major issues in research like the types of questions one can ask in a project, the role of time in research, and the different types of relationships one can estimate.
- Then one has to consider defining some basic terms like variable, hypothesis, data, and unit of analysis.
- A good research proposal gives you an opportunity to think through your project carefully, and clarify and define what you want to research.
- It provides you with an outline and to guide you through the research process.
- It also lets your supervisor and department or faculty know what you would like to research and how you plan to go about it.
- A hypothesis is a proposition – a tentative assumption which a researcher wants to test for its logical or empirical consequences. Hypotheses are more useful when stated in precise and clearly defined terms.

3.6 Keywords

Coding: Coding means assigning numbers to each of the answers, so that they can be analysed.

Data Collection: The search for answers to research questions is called data collection.

Editing: The data collected should be scanned to make sure that it is complete and that all the instructions are followed.

Problem Formulation: The problem formulation means converting the management problem to a research problem.

Research Process: Research process defines marketing research and describes the skills required to identify the problem, the decision alternatives, and the client's needs, which are critical components of a research activities.

Tabulation: It is the orderly arrangement data in a tabular form.

3.7 Review Questions

1. Research involves an eclectic blending of an enormous range of skills and activities. Comment.
2. Explain the concepts Nomothetic and Idiographic through examples only.
3. A white sales man sells less than black salesman. Is this hypothesis right? Why/why not?
4. A study was conducted to measure the motivation level of each of the category of managers. Formulate a hypothesis, suggesting testing procedures to show that there is no relation between the category of managers and the level of motivation.
5. Why should a hypothesis must be supported or backed up by theoretical relevance?
6. In your opinion, what might happen if the statistic under verification is not mentioned clearly?

Notes

Answers: Self Assessment

1. Probabilities
2. Longitudinal study
3. Relationship
4. Negative
5. Variable
6. Attribute
7. Extraneous
8. Null hypothesis
9. Choose a Test Criterion

3.8 Further Readings

Books

Cooper and Schinder, *Business Research Methods*, TMH.

OR Krishna Swamy, *Methodology of Research in Social Sciences*, HPH.

William MC Trochim, *Research Methods*, Biztantra.

William Zikmund, *Business Research Methods*, Thomson.

Online links

www.socialresearchmethods.net

www.experiment-resources.com

Unit 4: Research Problem

Notes

CONTENTS

Objectives
Introduction
4.1 Sources for Problem Identification
4.1.1 Self Questioning by Researcher while Defining the Problem
4.2 Selection of Problem
4.2.1 Selection Criteria
4.3 Understanding Problem
4.4 Necessity of Defined Problem
4.5 Pilot Testing
4.5.1 Data Collection
4.5.2 Data Processing
4.5.3 Analysis and Interpretation
4.6 Reporting the Results
4.7 Summary
4.8 Keywords
4.9 Review Questions
4.10 Further Readings

Objectives

After studying this unit, you will be able to:

- State the sources for problem identification;
- Identify the selection of problem;
- Define the concept of problem;
- Report the necessity of defined problem;
- Interpret the pilot testing;
- Summaries the reporting of results.

Introduction

There is a famous saying that "problem well-defined is half solved". This statement is strikingly true in market research, because if the problem is not stated properly, the objectives will not be clear. If the objective is not clearly defined, the data collection becomes meaningless.

The first step in research is to formulate the problem. A company manufacturing television sets might think that it is losing sales to a foreign company. A brief illustration aptly demonstrates how such problem can be ill-conceived. The management of a company felt, a drop in sales was because of the poor quality of product. Subsequently, research was undertaken with a view to

Notes

improve the quality of the product. But despite an improvement in quality, sales did not pick up. In this case, we may say that the problem is ill-defined. The actual reason was ineffective sales promotion. The problem thus needs to be carefully identified.

Caution It is vital and any error in defining the problem incorrectly can result in wastage of time and money.

Problem definition might refer to either a real-life situation or it may also refer to a set of opportunities. Market research problems or opportunities will arise under the following circumstances – (1) Unanticipated change (2) Planned change. Many factors in the environment can create problems or opportunities. Thus, changes in the demographics, technological and legal changes affect the marketing function. Now the question is how the company responds to new technology, or product introduced by the competitor or how to cope with the changes in lifestyles. It may be a problem and at the same time, it can also be viewed as an opportunity. In order to conduct research, the problem must be defined accurately.

While formulating the problem, clearly define:

1. Who is the focus?
2. What is the subject-matter of research?
3. To which geographical territory/area the problem refers to?
4. To which period does the study pertains to?

Example: "Why does the upper-middle class of Bangalore shop at Lifestyle during the Diwali season"?

Here all the above four aspects are covered. We may be interested in a number of variables due to which shopping is done at a particular place. The characteristic of interest to the researcher may be (1) Variety offered at Lifestyle (2) Discount offered by way of promotion (3) Ambience at the Lifestyle and the (4) Personalised service offered. In some cases, the cause of the problem is obvious whereas in others the cause is not so obvious. The obvious causes are the products being on the decline. Not so obvious causes could be a bad first experience for the customer.

4.1 Sources for Problem Identification

Research students can adopt the following ways to identify the problems:

- Research reports already published may be referred to define a specific problem.
- Assistance of any research organisation, which handles a number of projects of the companies, can be sought to identify the problem.
- Professors working in reputed academic institution can act as guides in problem identification.
- Company employees and competitors can assist in identifying the problems.
- Cultural and technological changes can act as a sources for research problem identification.
- Seminars/symposiums/focus groups can act as a useful source.

Notes

Notes Problem formulation is the key to research process. For a researcher, problem formulation means converting the management problem to a research problem. In order to attain clarity, the M.R manager and researcher must articulate clearly so that perfect understanding of each others is achieved. In research process, the first and foremost step happens to be that of selecting and properly defining a research problem. A researcher must find the problem and formulate it so that it becomes susceptible to research. Like a medical doctor, a researcher must examine all the symptoms (presented to him or observed by him) concerning a problem before he can diagnose correctly. To define a problem correctly, a researcher must know: what a problem is?

4.1.1 Self Questioning by Researcher while Defining the Problem

1. Is the research problem correctly defined?
2. Is the research problem solvable?
3. Can relevant data be gathered through the process of marketing research?
4. Is the research problem significant?
5. Can the research be conducted within the available resources?
6. Is the time given to complete the project sufficient?
7. What exactly will be the difficulties in conducting the study, and hurdles to be overcome?
8. Am I competent, to carry the study out?

Managers often want the results of research in accordance with their expectation. This satisfies them immensely. If one were to closely look at the questionnaire, it is found that in most cases, there are stereotyped answers given by the respondents. A researcher must be creative and should look at problems in a different perspective.

Task Cultural and technological changes can act as a source for research problem identification. Why/ why not?

Self Assessment

Fill in the blanks:

1. Any error in defining the problem incorrectly can result in wastage of ..... and .....
2. Managers often want the results of research in accordance with their .....
3. To define a problem correctly, a researcher must know: .....

4.2 Selection of Problem

The research problem undertaken for study must be carefully selected. The task is a difficult one, although it may not appear to be so. Help may be taken from a research guide in this connection. Nevertheless, every researcher must find out his own salvation for research problems cannot be borrowed. A problem must spring from the researcher's mind like a plant springing from its

Notes

own seed. If our eyes need glasses, it is not the optician alone who decides about the number of the lens we require. We have to see ourself and enable him to prescribe for us the right number by cooperating with him. Thus, a research guide can at the most only help a researcher choose a subject.

Inevitably, selecting a problem is somewhat arbitrary, idiosyncratic, and personal. Avoid selecting the first problem that you encounter. Try to select the most interesting and personally satisfying choice from among two or three possibilities. The problem selection should matter to you. You should be eager and enthusiastic.

Caution A good topic should be small enough for a conclusive investigation and large enough to yield interesting results. Remember that research must yield a publication for it to have meaning. You may wish to query likely periodical editors to see if they might be interested in an article on your research topic.

In some cases, as with a thesis or a dissertation, some sort of preliminary study may be needed to see if the problem and the study are feasible and to identify snags. Such a Pilot Study can be quite valuable.

Task Analyse what problems you might encounter while selecting a problem.

4.2.1 Selection Criteria

1. Your genuine enthusiasm for the problem.
2. Controversial subject should not become the choice of an average researcher.
3. The degree to which research on this problem benefits the profession and society.
4. The degree to which research on this problem will assist your professional goals and career objectives.
5. Too narrow or too vague problems should be avoided.
6. The degree to which this research will interest superiors and other leaders in the field.
7. The degree to which the research builds on your experience and knowledge.
8. Ease of access to the population to be studied and the likelihood that they will be cooperative Affordability.
9. Likelihood of publication.
10. Relationship to theories or accepted generalizations in the field.
11. Degree to which ethical problems are involved.
12. Degree to which research is unique or fills a notable gap in the literature.
13. Degree to which the research builds on and extends existing knowledge before the final selection of a problem is done, a researcher must ask himself the following questions:
    (a) Whether he is well equipped in terms of his background to carry out the research?
    (b) Whether the study falls within the budget he can afford?
    (c) Whether the necessary cooperation can be obtained from those who must participate in research as subjects?

Notes

Self Assessment

Fill in the blanks:

4. A good topic should be small enough for a .....investigation.
5. The research problem undertaken for study must be ..... selected.

4.3 Understanding Problem

Once the problem has been selected, the same has to be understood thoroughly and then the same has to be reframed into meaningful terms from an analytical point of view. The first step in research is to formulate the problem. A company manufacturing television sets might think that it is losing sales to a foreign company. A brief illustration aptly demonstrates how such problem can be ill-conceived. The management of a company felt, a drop in sales was because of the poor quality of product. Subsequently, research was undertaken with a view to improve the quality of the product. But despite an improvement in quality, sales did not pick up. In this case, we may say that the problem is ill-defined. The actual reason was ineffective sales promotion. The problem thus needs to be carefully identified. Marketing problem which needs research can be classified into two categories:

1. Difficulty related problems
2. Opportunity related problems, while the first category produces negative results such as, decline in market share or sales, the second category provides benefits.

Problem definition might refer to either a real-life situation or it may also refer to a set of opportunities. Market research problems or opportunities will arise under the following circumstances: (1) Unanticipated change (2) Planned change. Many factors in the environment can create problems or opportunities. Thus, changes in the demographics, technological and legal changes affect the marketing function. Now the question is how the company responds to new technology, or product introduced by the competitor or how to cope with the changes in life-styles. It may be a problem and at the same time, it can also be viewed as an opportunity. In order to conduct research, the problem must be defined accurately.

While formulating the problem, clearly define:

1. Who is the focus?
2. What is the subject-matter of research?
3. To which geographical territory/area the problem refers to?
4. To which period does the study pertains to?

Example: "Why does the upper-middle class of Bangalore shop at Life-style during the Diwali season"?

Here all the above four aspects are covered. We may be interested in a number of variables due to which shopping is done at a particular place. The characteristic of interest to the researcher may be (1) Variety offered at life-style (2) Discount offered by way of promotion (3) Ambience at the life-style and (4) Personalised service offered. In some cases, the cause of the problem is obvious whereas in others the cause is not so obvious. The obvious causes are the products being on the decline. Not so obvious causes could be a bad first experience for the customer.

4.4 Necessity of Defined Problem

Defining a research problem properly is a prerequisite for any study and is a step of the highest importance. A problem well defined is half solved. Defining the problem is often more essential than its solution because when the problem is formulated, an appropriate technique can be applied

Notes

to generate alternative solutions. This statement signifies the need for defining a research problem. The problem to be investigated must be defined unambiguously for that will help to discriminate relevant data from the irrelevant ones. A proper definition of research problem will enable the researcher to be on the track whereas an ill-defined problem may create hurdles. When you define a research problem you are trying to reduce the outcome of an answer. The question of course when you speak about "marketing research" is how I can target more customers that I can sell my product to. You are looking for specific answers such as: "What type of soda do all foreign born males between the ages of 25-35 drink?" This is defining the problem. What do you consider foreign born males? What constitutes soda? etc. This is important because companies and sales organization attempt to "target" their market instead of taking a shotgun approach. The process is to first make sure any information you obtain is credible and from a reputable organization. Then break down your problem and pick apart any inconsistencies you may see within you research project. Problem formulation is the key to research process. For a researcher, problem formulation means converting the management problem to a research problem.

In order to attain clarity, the manager and researcher must articulate clearly so that perfect understanding of each other's is achieved.

Self Assessment

Fill in the blanks:

6. In order to attain clarity, the manager and researcher must ..... clearly.
7. Problem ..... is the key to research process.
8. When you define a research problem you are trying to ..... the outcome of an answer.
9. Changes in the demographics, technological and legal changes affect the ..... function.

4.5 Pilot Testing

A pilot test is a method used to test the design and/or methods and/or instrument prior to carrying out the research. Basically, pilot testing means finding out if your survey, key informant interview guide or observation form will work in the "real world" by trying it out first on a few people. Can be small, 3-5, since the purpose is not to collect data but to refine your process and/or instrument.

Pilot testing involves conducting a preliminary test of data collection tools and procedures to identify and eliminate problems, allowing programs to make corrective changes or adjustments before actually collecting data from the target population. A pilot survey is very useful when the actual survey is to be on a big scale as it may provide data which will allow costs to be trimmed. Also, a pilot survey will give an estimate of the non-response rate and it will also give a guide as to the adequacy of the sampling frame chosen.

Did u know? **Why pilot test?**

The purpose is to make sure that everyone in your sample not only understands the questions, but understands them in the same way. This way, too, you can see if any questions make respondents feel uncomfortable. You'll also be able to find out how long it takes to complete the survey in real time.

Notes

A pilot test usually involves simulating the actual data collection process on a small scale to get feedback on whether or not the instruments are likely to work as expected in a "real world" situation. A typical pilot test involves administering instruments to a small group of individuals that has similar characteristics to the target population, and in a manner that simulates how data will be collected when the instruments are administered to the target population.

Pilot testing gives programs an opportunity to make revisions to instruments and data collection procedures to ensure that appropriate questions are being asked, the right data will be collected, and the data collection methods will work. Programs that neglect pilot testing run the risk of collecting useless data.

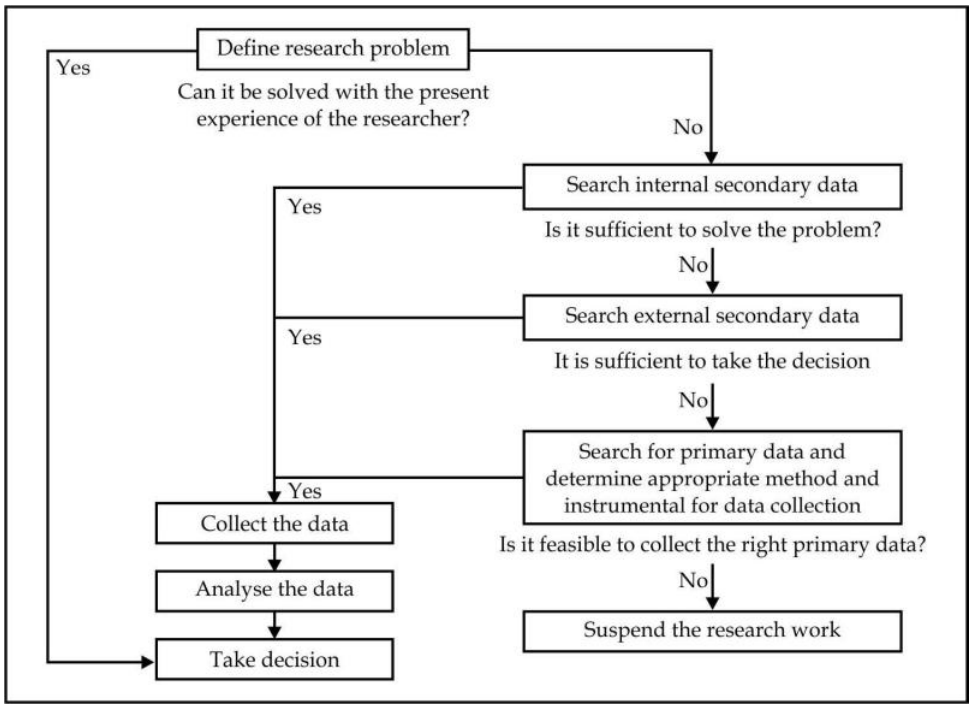
Pilot testing provides an opportunity to detect and remedy a wide range of potential problems with an instrument. These problems may include:

- Questions that respondents don't understand
- Ambiguous questions
- Questions that combine two or more issues in a single question (double-barreled questions)
- Questions that make respondents uncomfortable

Pilot testing can also help programs identify ways to improve how an instrument is administered. For example, if respondents show fatigue while completing the instrument, then the program should look for ways to shorten the instrument. If respondents are confused about how to return the completed instrument, then the program needs to clarify instructions and simplify this process.

4.5.1 Data Collection

Data collection is the systematic recording of information; data analysis involves working to uncover patterns and trends in data sets; data interpretation involves explaining those patterns and trends. Scientists interpret data based on their background knowledge and experience, thus different scientists can interpret the same data in different ways.



Notes

4.5.2 Data Processing

Processing data is very important in market research. After collecting the data, the next job of the researcher is to analyze and interpret the data. The purpose of analysis is to draw conclusion. There are two parts in processing the data.

1. Data Analysis
2. Interpretation of data

Analysis of the data involves organizing the data in a particular manner. Interpretation of data is a method for deriving conclusions from the data analyzed. Analysis of data is not complete, unless it is interpreted.

Steps in Processing of Data

1. Preparing raw data
2. Coding
3. Editing
4. Tabulation of data
5. Summarising the data
6. Usage of statistical tool.

Preparing Raw Data

Data collection is a significant part of market research. Even more significant is, to filter out the relevant data from the mass of data collected. Data continues to be in raw form, unless they are processed and analyzed.

Primary data collected by surveys, observations by field investigations are hastily entered into questionnaires. Due to the pressure of interviewing, the researcher has to write down the responses immediately. Many times this may not be systematic. The information so collected by field staff is called raw data.

The information collected may be illegible, incomplete and inaccurate to some extent. Also the information collected will be scattered in several data collection formats. The data lying in such a crude form are not ready for analysis. Keeping this in mind the researcher must take some measures to organize the data, so that it can be analyzed.

The various steps which are required to be taken for his purpose are (a) editing and (b) coding and (c) tabulating.

Coding

Coding refers to all those activities which helps in transforming edited questionnaires into a form which is ready for analysis. Coding speeds up the tabulation while editing eliminates errors. Coding involves assigning numbers or other symbols to answers, so that the responses can be grouped into limited number of classes or categories.

Example: 1 is used for male and 2 for female.

Notes

Editing

The main purpose of editing is to eliminate errors and confusion. Editing involves inspection and correction of each questionnaire. The main role of editing is to identify commissions, ambiguities and errors in response.

Therefore editing means, the activity of inspecting, correcting and modifying the correct data.

Tabulation of Data

Tabulation refers to counting the number of cases that fall into various categories. The results are summarized in the form of statistical tables. The raw data is divided into groups and subgroups. The counting and placing of data in particular group and subgroup are done. Tabulation involves

1. Sorting and counting
2. Summarizing of data

Tabulation may be of two types (1) simple tabulation (2) cross tabulation. In simple tabulation, a single variable is counted. Cross tabulation includes 2 or more variables, which are treated simultaneously. Tabulation can be done entirely by hand or by machine or both hand and machine.

Summarising the Data

Before taking up summarizing, the data should be classified into (1) Relevant data (2) Irrelevant data. During the field study, the researcher has collected lot of data which he may think would be of use. Summarizing the data includes (1) Classification of data (2) Frequency distribution (3) Use of appropriate statistical tool.

Classification of Data

(a) ***Number of Groups:*** Number of groups should be sufficient to record all possible data. Classification should not be too narrow. If it is too narrow, there can be an overlap.

Example: If a researcher is conducting a survey on "Why the current car owner dislikes the car"? The car owner may indicate the following:

1. Difficulty in seeking entry to the back seat
2. Interior space
3. Cramped leg room
4. Mileage
5. Rattling of the engine
6. Dickey space

Now all the above data can be classified into 2 or 3 categories such as (1) Discomfort (2) Expense (3) Pride (4) Safety (5) Design of the car.

(b) ***Width of the Class Interval:*** Class interval should be uniform and should be of equal width. This will give consistency in the data distribution.

(c) ***Exclusive Categories:*** Classification made should be done in such a way that, the response can be placed in only one category.

Notes

Example: Problem of Leg room is the answer by respondent. This should be placed either under Discomfort or Design but not both.

(d) ***Exhaustive Categories:*** This should be made to include all responses including "Don't Know" answers. Sometimes this will influence the ultimate answer to the research problem.

(e) ***Avoid Extremes:*** Avoid open ended class interval.

Usage of Statistical Tools

Frequency Distribution: Frequency distribution, simply reports the number of responses that each question received. Frequency distribution, organizes data into classes or groups. It shows the number of data that falls into particular class.

Example: Frequency distribution:

Income	No. of people
4000-6999	100
7000-9999	122
10000-12999	140

In marketing research central value or tendency plays a very important role. The researcher may be interested in knowing the average sales/shop, average consumption per month etc. The population parameters can be calculated with the help of simple average. The average of sample may be taken as population parameter. E.g. If the average income of the population is to be computed, the researcher may select a sample, collect data on family income and calculate the relevant statistics which will be a representative of the population.

The total purchasing power of the community can be estimated on sample average. If the sample is stratified, the purchasing power of each income class may also be estimated. The median figure will reveal that half the population has more income than the median income, and half the population has less income than median income. The mode will reveal the most common frequency. Based on this, shoppers can play their strategy to sell the product.

The three most common ways to measure centrality or central tendency is mode, median and mean.

Mode

The mode is the central value or item, that occurs most often, when data is categorized in a frequency distribution, it is very easy to identify the mode, since the category in which the mode lies has the greatest number of observations.

Example: Data regarding household income of 300 people as tabulated by researcher.

Income (₹)	Number (f)	Cumulative Frequency
upto 10000	30	30
10000-14999	125	155
20000-24999	50	205
25000-29999	30	235

Contd...

Notes

30000-34999	33	268
35000-49999	20	288
above 35000	12	300

In the above Table 125 is the modal class.

Mode can be calculated using the formula:

$$M_0 = LLM_0\left[\frac{D_1}{D_1 + D_2}\right] \times i$$

LM_0 = Lower limit of modal class

D_1 = Difference between the frequency of modal class and the class immediately preceding the modal class

D_2 = Difference between the frequency of the modal class and the class immediately succeeding the modal class.

i = size of the modal class interval

$$M_d = 10,000 + \left(\frac{95}{95+75}\right) \times 5,000$$

substitute the values

$$= 10000 + \left(\frac{95}{170}\right) = 5000 = 1000+2794 = ₹\ 12794$$

Conclusion: Majority have the income of ₹ 12794. This is how statistical techniques are used in MR application.

Median

Median lies precisely halfway between highest and lowest values. It is necessary to arrange the data into ascending or descending order before selecting the median value. For ungrouped data with an odd number of observation, the median would be the middle value. For even number of observations, the median value is half way between central value.

For a grouped data median is calculated using the formula

$$M_d = LM_d \frac{\left(\frac{N}{2} - C.F.\right)}{FM_d} \times i$$

M_d = Lower limit of median class

CF = Cumulative frequency for the class just below the median class.

Fm_d = Frequency of the median class.

i = Size of the class interval of median class.

In the table N = 300 N/2 = 150. The class containing the 150th person is the median class.

Substitute the value, we get median M_d = 21568

Notes

Conclusion: Half of the population has income> 21568' and half of the population has income < 21568.

Mean

In a grouped data, the midpoint of each category would be multiplied by the number of observation in that category. Sum up and the total to be divided by the total number of observation.

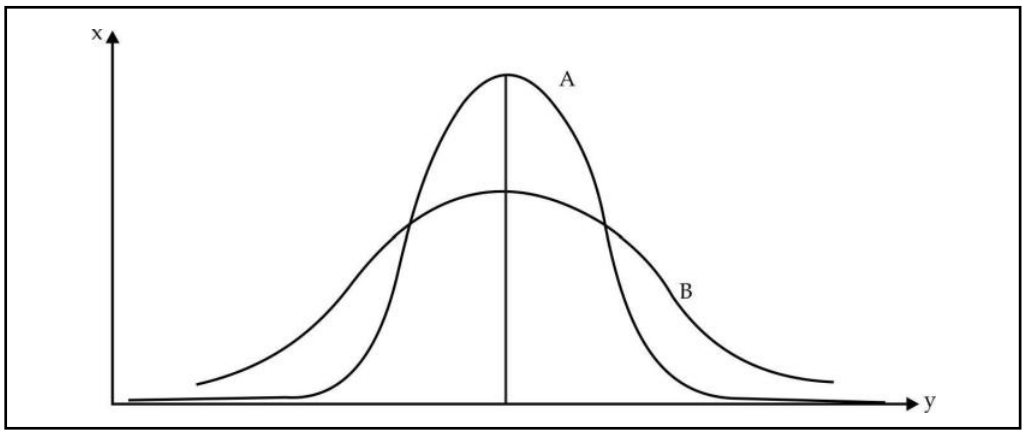
$$\text{Eqn., } X = \Sigma\left(\frac{fx}{\Sigma f}\right)$$

Example: 2 students X, Y attend 3 classes tests and the scores areas follows:

Though Mean is same, X is better than Y.

Measures of Dispersion

Marks		1st Test	2nd Test	3rd Test	Mean	
X		55%	60%	65%	60%	
Y		65%	60%	55%	60%	
Conclusion	X	- has improved				
	Y	- has Deteriorated				



Dispersion is the spread of the data in a distribution. A measure of dispersion indicates the degrees of scattered ness of the observations. Let curves A and B represent two frequency distributions. Observe that A and B have the same mean. But curve A has less variability than B.

If we measure only the mean of these two distributions, we will miss an important difference between A and B. To increase our understanding of the pattern of the data we must also measure its dispersion.

Range: It is the difference between the highest and lowest observed values.

i.e. range = H – L, H = Highest, L = Lowest.

Note:

1. Range is the crudest measure of dispersion.
2. $\frac{H-L}{H+L}$ is called the coefficient of range.

Notes

Semi-inter quartile range (Quartile deviation) semi-inter quartile range Q.

Q is given by $Q = \frac{Q_3 - Q_1}{2}$

Note:

1. $\frac{Q_3 - Q_1}{Q_3 + Q_1}$ is called the coefficient of quartile deviation.

2. Quartile deviation is not a true measure of dispersion but only a distance of scale.

Mean Deviation (MD): If A is any average then mean deviation about A is given by

$$MD(A) = \frac{\sum f_i |x_i - A|}{N}$$

Note:

1. Mean deviation about mean MD $(\bar{x}) = \frac{\sum f_i |x_i - \bar{x}|}{N}$

2. Of all the mean deviations taken about different averages mean derivation about the median is the least.

3. $\frac{MD(A)}{A}$ is called the coefficient of mean deviation.

Variance and Standard Deviation

Variance (σ^2): A measure of the average squared distance between the mean and each term in the population.

$$\sigma^2 = \frac{1}{N}\sum f_i (x_i - \bar{x})^2$$

Standard deviation (σ) is the positive square root of the variance

$$\sigma = \sqrt{\frac{1}{N}\Sigma f_i (x_i - \bar{x})^2}$$

$$\sigma^2 = \frac{1}{N}\sum f_i (x_i^2 - (\bar{x})^2$$

Note: Combined variance of two sets of data of N_1 and N_2 items with means x_1 and x_2 and standard deviations σ_1 and σ_2 respectively is obtained by

$$\sigma^2 = \frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_1 d_2^2}{N_1 + N_2}$$

Where, $d_1^2 = (\bar{x} - \bar{x}_1)^2,\ d_2^2 = (\bar{x} - \bar{x}_2)^2$

Notes

and $$\bar{x} = \frac{N_1\overline{x_1} + N_2\overline{x_2}}{N_1 + N_2}$$

Sample variance **(σ^2):** Let $x_1, x_2, x_3, \ldots\ldots\ldots x_n$, represents a sample with mean $\bar{x}$

Then sample variance σ^2 is given by

$$\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

$$= \frac{\sum x^2}{n-1} - \frac{n(\bar{x})^2}{n-1}$$

Note: $\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{\Sigma x^2}{n-1} - \frac{n(\bar{x})^2}{n-1}}$ is called the sample standard deviation.

Coefficient of Variation (C.V.)

It is a relative measure of dispersion that enables us to compare two distributions. It relates the standard deviation and the mean by expressing the standard deviation as a percentage of the mean.

$$\text{C.V.} = \frac{\sigma}{\bar{x}} \times 100$$

Note:

1. Coefficient of variation is independent of the unit of the observation.
2. This measure cannot be used when x is zero or close to zero.

Illustration 1: For the data 103, 50, 68, 110, 105, 108, 174, 103, 150, 200, 225, 350, 103 find the Range, Coefficient of range and coefficient of quartile deviation.

Solution: Range = H – L = 350 – 50 = 300

$$\text{Coefficient of range} = \frac{H-L}{H+L} = \frac{300}{350+50} = 0.7$$

To find Q_1 and Q_3 we arrange the data in ascending order

50, 68, 103, 103, 103, 103, 105, 108, 110, 150, 174, 200, 225, 350,

$$\frac{n+1}{4} = \frac{14}{4} = 3.5$$

$$\frac{3(n+1)}{4} = 10.5$$

$\therefore$ $Q_1 = 103 + 0.5\,(103 - 103) = 103$

$Q_3 = 174 + 0.5\,(200 - 174) = 187$

$$\text{Coefficient of QD} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Notes

$$= \frac{84}{290} = 0.2896$$

Illustration 2: Calculate coefficient of mean deviation about (i) Median (ii) mean from the following data

X	14	16	18	20	22	24	26
f	2	4	5	3	2	1	4

Solution:

X	F	Cf	fx	$\lvert x - \bar{x} \rvert$	$\lvert x - M \rvert$	$f \lvert x - \bar{x} \rvert$	$f \lvert x - M \rvert$
14	2	2	28	5.71	4	11.42	8
16	4	6	64	3.71	2	14.84	8
18	5	11	90	1.71	0	8.55	0
20	3	14	60	0.29	2	0.87	6
22	2	16	44	2.29	4	4.58	8
24	1	17	24	4.29	6	4.29	6
26	4	21	104	6.29	8	25.16	32
	21		**414**			**69.71**	**68**

$$\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{414}{21} = 19.71$$

$$\frac{N+1}{2} = \frac{22}{2} = 11 \therefore \text{Median } M = 18$$

$$\text{Now (i) M.D. } (\bar{x}) = \frac{\sum f_i \lvert x_i - \bar{x} \rvert}{N} = \frac{69.71}{21} = 3.32$$

$$\text{Coefficient of MD}(\bar{x}) = \frac{MD(\bar{x})}{\bar{x}} = \frac{3.32}{19.71} = 0.16$$

$$\text{(ii)} \qquad \text{M.D. (M)} = \frac{\sum f_i \lvert x_i - M \rvert}{N} = \frac{68}{21} = 3.24$$

$$\text{Coefficient of MD (M)} = \frac{MD(M)}{M} = \frac{3.24}{18} = 0.18$$

Illustration 3: A purchasing agent obtained a sample of incandescent lamps from two suppliers. He had the sample tested in his laboratory for length of life with following results.

Length of Light in hours	Sample A	Sample B
700 – 900	10	3
900 – 1100	16	42
1100 – 1300	26	12
1300 – 1500	8	3

Which company's lamps are more uniform.

Notes

Solution:

Class interval	Sample A	Midpoint x	$u = \frac{x-1000}{200}$	Fu	fu^2
700 – 900	10	800	– 1	– 10	10
900 – 1100	16	1000	0	0	0
1100 – 1300	26	1200	1	26	26
1300 – 1500	8	1400	2	16	32
	60			32	68

$$\bar{u}_A = \frac{32}{60} = 0.533$$

$$\bar{x}_A = 1000 + 200$$

$$\therefore \quad \bar{x}_A = 1000 + 200\,(0.533) = 1106.67$$

$$\sigma_u^2 = \frac{1}{N} = \Sigma fu^2 - (\bar{u}) = \frac{68}{60} - (0.533)^2$$

$$= 1.133 - 0.2809$$

$$\sigma_u^2 = 0.8524$$

$$\sigma_u = 0.9233$$

$$\sigma_x = 200 \times 0.9233 = 184.66$$

$$\therefore \quad \text{CV for sample A} = \frac{\sigma_A}{\bar{x}_A} \times 100$$

$$= \frac{184.66}{1106.67} \times 100 = 16.68\,\%$$

Class interval	Sample B	Midpoint x	$u = \frac{x-1000}{200}$	fu	fu^2
700 – 900	3	800	– 1	– 3	3
900 – 1100	42	1000	0	0	0
1100 – 1300	12	1200	1	12	12
1300 – 1500	3	1400	2	6	12
	60			15	27

$$\bar{V} = \frac{15}{60} = 0.25$$

$$\bar{x}_B = 1000 + 200\,\bar{V}$$

$$= 1000 + 58$$

$$\therefore \quad \bar{x}_B = 1058$$

$$\sigma_v^2 = \frac{1}{N}\Sigma fv^2 - (\bar{V})^2 = \frac{27}{60} - (0.25)^2$$

$$= 0.45 - 0.0625$$

Notes

$$\sigma_v^2 = 0.3875$$

$$\sigma_v = 0.6225$$

$$\sigma_B = 200\ s_v$$

$$= 200 \times 0.6225 = 124.5$$

$$\text{C.V for Sample B} = \frac{\sigma_B}{\bar{x}_B} \times 100$$

$$\frac{124.5}{1058} \times 100 = 11.77\%$$

Since C.V. for sample B is smaller, sample B lamps are more uniform.

4.5.3 Analysis and Interpretation

Interpretation means bring out the meaning of data or we can say that interpretation is to convert data into information. The essence of any research is to draw conclusion about the study. This requires high degree of skill. There are two methods of drawing conclusions (1) induction (2) deduction.

In induction method, one starts from observed data and then generalization is done, which explains the relationship between objects observed.

On the other hand, deductive reasoning starts from some general law and then applied to a particular instance i.e., deduction comes from general to a particular situation.

Example of induction: All products manufactured by Sony are excellent. DVD player model 2602MX is made by Sony. Therefore it must be excellent.

Example of deduction: All products have to reach decline stage one day and become obsolete. This Radio is in decline mode. Therefore it will become obsolescent.

During inductive phase, we reason from observation. During deductive phase, we reason towards observation. Both logic and observation are essential for interpretation.

Successful interpretation depends on 'How Well the data is analyzed'. If data is not properly analyzed, the interpretation may go wrong. If analysis has to be corrected, then data collection must be proper. Similarly if data collected is proper but analyzed wrongly, then also the interpretation or conclusion will be wrong. Sometimes even with proper data and proper analysis, can still lead to wrong interpretation. Interpretation depends on. Experience of the researcher and methods used by him for interpretation.

Example: A detergent manufacturer is trying to decide, which of the three sale promotion methods (Discount, contest, buy one get one free) would be most effective in increasing the sales. Each sales promotion method is run at different times in different cities. The sales got by the different sale promotion is a follows.

Sales Impact of Different Sale Promotion Methods

Sales Promotion Method	Sales Associated with Sales Promotion
1	2000
2	3500
3	2510

Notes

The results can conclude that the second Sales Promotion method was the most effective in developing sales. This may be adopted nationally to promote the product. But one cannot say that the same method of sales promotion will be effective in each and every city under study.

Precautions to be taken while Interpreting the Marketing Research Data

1. Keep the main objective of the research in mind.
2. Analysis of data should start from simpler and more fundamental aspects.
3. It should not be confusing.
4. Sample size should be adequate.
5. Take care before generalization of the sample studied.
6. Give due attention to significant questions.
7. Do not miss the significance of some answers, because they are found from a very few respondents, such as "don't know" or "can't say".

4.6 Reporting the Results

The goal of research is not just to discover something but to communicate that discovery to a larger audience—other social scientists, government officials, your teachers, the general public—perhaps several of these audiences. Whatever the study's particular outcome, if the research report enables the intended audience to comprehend the results and learn from them, the research can be judged a success. If the intended audience is not able to learn about the study's results, the research should be judged a failure no matter how expensive the research, how sophisticated its design, or how much of yourself you invested in it.

This conclusion may seem obvious, and perhaps a bit unnecessary. After all, you may think that all researchers write up their results for other people to read. But the fact is that many research projects fail to produce a research report. Sometimes the problem is that the research is poorly designed to begin with and cannot be carried out in a satisfactory manner; sometimes unanticipated difficulties derail a viable project. But too often the researcher just never gets around to writing a report. And then there are many research reports that are very incomplete or poorly written or that speak to only one of several interested audiences. The failure may not be complete, but the project's full potential is not achieved.

The stage of reporting research results is also the point at which the need for new research is identified. It is the time when, so to speak, "the rubber hits the road"—when we have to make our research make sense to others. To whom will our research be addressed? How should we present our results to them? Should we seek to influence how our research report is used?

The research report will present research findings and interpretations in a way that reflects some combination of the researcher's goals, the research sponsor's goals, the concerns of the research subjects, and perhaps the concerns of a wider anticipated readership. Understanding the goals of these different groups will help the researcher begin to shape the final report even at the start of the research. In designing a proposal and in negotiating access to a setting for the research, commitments often must be made to produce a particular type of report, or at least cover certain issues in the final report. As the research progresses, feedback about the research from its participants, sponsoring agencies, collaborators, or other interested parties may suggest the importance of focusing on particular issues in the final report. Social researchers traditionally have tried to distance themselves from the concerns of such interested parties, paying attention

Notes

only to what is needed to advance scientific knowledge. But in recent years, some social scientists have recommended bringing these interested parties into the research and reporting process itself.

Self Assessment

Fill in the blanks:

10. A pilot survey is very useful when the actual survey is to be on a ..... .
11. Coding speeds up the tabulation while editing eliminates ..... .
12. ..... refers to counting the number of cases that fall into various categories.
13. ..... lies precisely halfway between highest and lowest values.
14. Coefficient of variation (C.V.) is a relative measure of dispersion that enables us to compare ..... .

4.7 Summary

- Proper problem formulation is the key to success in research.
- It is vital and any error in defining the problem incorrectly can result in wastage of time and money.
- Several elements of introspection will help in defining the problem correctly.
- The task of defining a research problem, very often, follows a sequential pattern.
- The problem is stated in a general way, the ambiguities are resolved, thinking and rethinking process results in a more specific formulation of the problem.
- It is done so that it may be a realistic one in terms of the available data and resources and is also analytically meaningful.
- All this results in a well defined research problem that is not only meaningful from an operational point of view.
    - But is equally capable of paving the way for the development of working hypotheses and for means of solving the problem itself.
    - Data when collected is raw in nature. When processed, it becomes information without data analysis, and interpretation, researcher cannot draw any conclusion.
    - Interpretation can use either induction or deduction logic. While interpreting certain precautions are to be taken.

4.8 Keywords

Data collection: Data collection is the systematic recording of information; data analysis involves working to uncover patterns and trends in data sets; data interpretation involves explaining those patterns and trends.

Editing: Editing, include inspection and correction of each questionnaire.

Frequency Distribution: Frequency distribution, organizes data into classes or groups.

Median: Median lies precisely halfway between highest and lowest values.

Notes

Mode: The mode is the central value or item that occurs most often, when data is categorized in a frequency distribution.

Objective of Research: It means to what the researcher aims to achieve.

Pilot Study: A small scale preliminary study conducted before the main research in order to check the feasibility or to improve the design of the research.

Research Problem: It focuses on the relevance of the present research.

4.9 Review Questions

1. The objective of research problem should be clearly defined; otherwise the data collection becomes meaningless. Discuss with suitable examples.
2. Cultural and technological changes can act as a source for research problem identification. Why/why not?
3. Defining a research problem properly is a prerequisite for any study. Why?
4. What precautions should be taken while formulating a problem?
5. If you are appointed to do a research for some problem with the client, what would you take as the sources for problem identification?
6. It may be a problem and at the same time, it can also be viewed as an opportunity. Why/ why not?
7. In some cases, some sort of preliminary study may be needed. Which cases are being referred to and why?
8. A problem well defined is half solved. Comment.
9. Explain the following:
   (a) Mode
   (b) Median
   (c) Mean
10. How to summarise and classify the collected data?

Answers: Self Assessment

1. Time, Money
2. Expectation
3. What a problem is
4. Conclusive
5. Carefully
6. Articulate
7. Formulation
8. Reduce
9. Marketing
10. Big scale

Notes

11. errors
12. Tabulation
13. Median
14. Two distributions

4.10 Further Readings

Books

C R Kotari, *Research Methodology*, Vishwa Prakashan.

Cooper and Schinder, *Business Research Methods*, TMH.

David Luck and Ronald Rubin, *Marketing Research*, PHI.

Naresh Amphora, *Marketing Research*, Pearson Education.

S. N. Murthy and U. Bhojanna, *Business Research Methods*, 2nd Edition, Excel Books.

Online links

www.experiment-resources.com

www.scribd.com

Notes

Unit 5: Review of Literature in Research

CONTENTS

Objectives
Introduction
5.1 Use of Literature Review
5.2 Search for Related Literature
5.3 Reading the Literature
5.4 Guidelines for Information Presentation
5.5 Process of Literature Review
5.6 Summary
5.7 Keywords
5.8 Review Questions
5.9 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss the use of literature review
- Discuss the reading of the literature
- State the guidelines for information presentation
- Recognize the process of literature review

Introduction

A literature review is an evaluative report of information found in the literature related to your selected area of study. The review should describe, summarise, evaluate and clarify this literature. It should provide a theoretical base for the research and help you (the author) determine the nature of your research. Works which are irrelevant should be discarded and those which are peripheral should be looked at critically.

In writing the literature review, the purpose is to convey to the reader what knowledge and ideas have been established on a topic, and what their strengths and weaknesses are. As a piece of writing, the literature review must be defined by a guiding concept.

Example: Research objective, the problem or issue being discussed.

It is not just a descriptive list of the material available, or a set of summaries.

Notes Besides enlarging the body of knowledge about the topic, writing a literature review leads the writer to gain and demonstrate skills in the following areas:

1. ***Searching skills:*** It improves the ability of the researcher to sift the literature efficiently, using manual or computerized methods, to identify a set of useful articles and books.

Contd...

Notes

2. ***Analysing skills:*** It is the ability to apply principles of analysis to identify the unbiased and valid studies?
3. ***Application of new approaches and methods:*** It helps the researcher to understand new approaches and methods to deal with the research problems.

5.1 Use of Literature Review

As researcher, it is important to understand the purpose of investigating the literature related to the research project. The literature review assists the researcher in knowing the ways and means to deal with the research problem.

- It helps the researcher to learn about the studies similar to his own study and the research design and methodology adopted to carry out those studies by earlier researchers.
- It provides useful source of data related to the subject being studied.
- It helps in introducing important and useful research personalities.
- It provides an opportunity to see the study in a historical perspective.
- Literature review provides new ideas, methods and approaches to deal with research problems.
- It helps the researcher to compare his own study with other relevant studies.
- It helps in anticipating the problems arising during the collection of data. The researcher can therefore take precautions to overcome those problems.

Did u know? **Objectives of Literature Review**

Information in literature review should be organised and related directly to the research problem.

Task Elucidate upon the purpose of writing literature review.

Self Assessment

Fill in the blanks:

1. A ..... is an evaluative report of information found in the literature related to your selected area of study.
2. The literature review assists the researcher in knowing the ways and means to deal with the …….

5.2 Search for Related Literature

When beginning a search for related literature, practical research suggests travelling to the library and looking at the selection of indices, abstracts and available bibliographies. Information available on microfilm must also be considered.

World Wide Web (internet) a boon for researchers help in identifying useful and relevant data to the research project.

Notes

Association of Indian Universities periodical *University News* publishes the thesis of the month in the last page of the periodical. It includes the research projects completed in that particular month, and is an important source of literature review.

Research agencies conduct various studies which comprises abundant data that may be helpful in searching the literature.

5.3 Reading the Literature

1. Record the problem at the top of a sheet of paper.
2. Record each sub-problem in full, serially, across the page.
3. Study each sub-problem, separating out the keywords.
4. Record the keywords or phrases in a column under the sub-problem.
5. Consult the index, bibliographies and abstracts to find books, articles etc., armed with the "identified keywords" of your problem.
6. Read! Read! Read!

5.4 Guidelines for Information Presentation

1. ***Discuss fairly and clearly:*** The review of literature should be like a discussion with a friend concerning the studies, research reports, and writings that bear directly on your own efforts. Be very clear in your thinking.
2. ***Organize a plan:*** Begin the discussion from a broad perspective and narrow it down to the specific problem of the research.
3. ***Do not copy information as it is:*** Emphasise over what the study interprets rather to what content has been produced. So, the researcher should critically evaluate and present the information in his own words.
4. ***Establish the relationship between literature and research project:*** This can be done by charting each study in relation to the problem or sub- problem it addresses. Study carefully before beginning to write. Literature discussed should have a link to the research problem.
5. ***Summarise:*** Summarise major contributions of significant studies and articles to the body of knowledge under review, maintaining the focus established in the introduction.

Caution Do not forget to evaluate the current "state of the art" for the body of knowledge reviewed, pointing out major methodological flaws or gaps in research, inconsistencies in theory and findings, and areas or issues pertinent to any future study.

Conclude by providing some insight into the relationship between the central topic of the literature review and a larger area of study such as a discipline, a scientific endeavour, or a profession.

Self Assessment

Fill in the blanks:

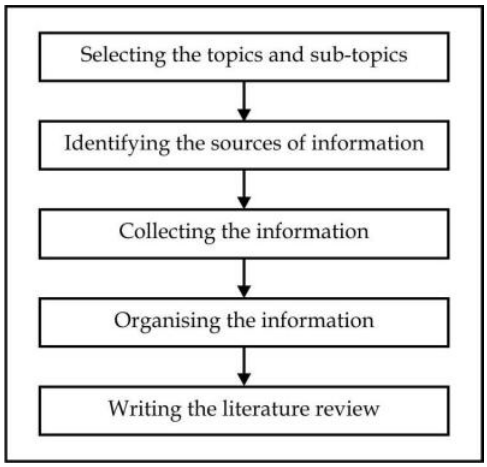
3. One should always ..... major contributions of significant studies.
4. When ..... a search for related literature, practical research suggests travelling to the library.

5.5 Process of Literature Review

Notes

Review of literature is a process of systematic selection, evaluation, correlation and present information relevant to a research question being studied. There are five steps in writing a literature review.

1. ***Selecting the topics and sub-topics:*** The researcher needs to select the topics and sub-topics related to the research question being studied. It helps to direct the literature search in the right direction. If the topics and sub-topics are not chosen, the researcher may end up with lot of unrelated information.
2. ***Identifying the sources of information:*** The sources of information for specific topics are to be identified and a list of the sources along with the specific topic should be made.
3. ***Collecting the information:*** Collect information systematically one after the other from the reliable sources. The information about each topic should be recorded separately. This helps the researcher to organise the information properly.



4. ***Organise the information:*** Information about each topic should be recorded and maintained separately. It has to be categorised based on the topics and sub-topics. The categorised information may be used appropriately in writing a literature review.
5. ***Writing the literature review:*** It consists of three steps, which are explained below:

Did u know? **What are the steps involved in literature review writing?**

Introduction: Define the topic, issues or areas of research being studied, thus providing an appropriate context for reviewing the literature.

Body: Critically evaluate the information and make appropriate comparison of the studies reviewed.

Conclusion: State what is your view point about the study, but not what the study says.

Self Assessment

Fill in the blanks:

5. The researcher needs to select the topics and sub-topics related to the ..... being studied.

Notes

6. If the ..... and ..... are not chosen, the researcher may end up with lot of unrelated information.
7. The sources of information for specific topics are to be identified and a list of the sources along with the .....should be made.
8. Introduction primarily involves ..... the topic, issues or areas of research being studied.

5.6 Summary

- Market researcher can avail wealth of data from secondary data.
- There are many benefits that accrue from literature review.
- Literature review provides useful data which aids the researcher for better focus.
- Literature review helps the researcher to move from management question to research question.
- Literature research enable to gather data for the current project which otherwise has been collected for some other purpose.
- This if found useful for current study will save time and cost.
- It is advised always, to refer to original source.
- Examine and scrutinise the relevance of data.
- Sometimes the secondary data through literature review may not be entirely suitable; however it can be a useful pointer, on how to design a good research study.

5.7 Keywords

Analysing skills: It is the ability to apply principles of analysis to identify the unbiased and valid studies.

Literature review: A literature review is an evaluative report of information found in the literature related to your selected area of study.

Research: Research is an art of scientific investigation.

5.8 Review Questions

1. In writing the literature review, the purpose is to convey to the reader what knowledge and ideas have been established on a topic, and what their strengths and weaknesses are?
2. What specific things would you keep in mind while writing the literature review?
3. What would happen if fail to choose topic and sub topic?
4. In your opinion, why is it suggested to begin a discussion from a broad perspective?
5. Is there any need to record the keywords or phrases? Why/why not?
6. Why is said that the review of literature should be like a discussion with a friend?
7. Do you think that a researcher must study each sub problem separately? Why/why not?
8. Why would you say world wide web to be a boon for researchers?

Notes

9. Why is said that a researcher should never copy the information as it is?
10. What might happen if in the conclusion, the researcher states that what study says?

Answers: Self Assessment

1. Literature review
2. Research problem
3. Summarise
4. Beginning
5. Research question
6. Topics, Sub-topics
7. Specific topic
8. Defining

5.9 Further Readings

Books

Abrams, M.A., *Social Surveys and Social Action*, London: Heinemann, 1951.

Arthur, Maurice, *Philosophy of Scientific Investigation* , Baltimore: John Hopkins University Press, 1943.

Bernal, J.D., *The Social Function of Science*, London: George Routledge and Sons, 1939.

Chase, Stuart, *The Proper Study of Mankind: An inquiry into the Science of Human Relations*, New York, Harper and Row Publishers, 1958.

S.N Murthy and U. Bhojanna, *Business Research Methods*, 2nd Edition, Excel Books.

Online links

ibrary.ucsc.edu

planningcommission.nic.in

Notes

Unit 6: Research Design

CONTENTS

Objectives
Introduction
6.1 Meaning
6.2 Types of Research Designs
6.3 Exploratory Research
6.3.1 Exploratory Research Methods
6.4 Conclusive Research
6.4.1 Descriptive or Diagnostic Research
6.4.2 Survey
6.4.3 Observation Studies
6.5 Causal Research
6.6 Experimentation
6.6.1 Test Units
6.6.2 Explanatory Variable
6.6.3 Dependent Variable
6.6.4 Extraneous Variables
6.7 Experimental Designs
6.7.1 Purely Post-design
6.7.2 Before-After Design
6.7.3 Factorial Design
6.7.4 Latin Square Design
6.7.5 Ex-post Facto Design
6.8 Summary
6.9 Keywords
6.10 Review Questions
6.11 Further Readings

Objectives

After studying this unit, you will be able to:

- Construct an overview of research design;
- Define exploratory research design;
- Explain the methods that are adopted during exploratory research;
- Describe the descriptive research design;
- Discus the causal research design;
- Explain the experimentation.

Notes

Introduction

Suppose a manufacturer of a quality machine finds sales disappointing and believes that they may be helped by the development of point of purchase display. The contemplated display is expensive but the manufacturer would like to try it out first on a limited basis to be sure that it stimulates more sales and profits than it costs. This can be achieved through a better planning and formulation of a good strategy. According to Kerlinger, "Research design is the plan, structure and strategy of investigation conceived so as to obtain answers to research questions and to control variance".

Research design is in fact the conceptual structure within which the research is conducted. Bernard Philips has described the research design as a "blue print for the collection, measurement and analysis of data".

6.1 Meaning

Research design is simply a plan for a study. This is used as a guide in collecting and analyzing the data. It can be called a blue print to carry out the study. It is like a plan made by an architect to build the house, if a research is conducted without a blue print, the result is likely to be different from what is expected at the start. The blue print includes (1) interviews to be conducted, observations to be made, experiments to be conducted data analysis to be made. (2) Tools used to collect the data such as questionnaire (3) what is the sampling methods used.

Research design can be thought of as the structure of research – it is the "glue" that holds all of the elements in a research project together. A successful design stems from a collaborative process involving good planning and communication.

Research Design is mainly of three types namely, exploratory, descriptive and causal research.

Exploratory research is used to seek insights into general nature of the problem. It provides the relevant variable that need to be considered. In this type of research, there is no previous knowledge; research methods are flexible, qualitative and unstructured.

Notes The researcher in this method does not know "what he will find".

Descriptive research is a type of research, very widely used in marketing research. Generally in descriptive study there will be a hypothesis, with respect to this hypothesis, we ask questions like size, distribution, etc.

Causal research, this type of research is concerned with finding cause and effect relationship. Normally experiments are conducted in this type of research.

Self Assessment

Fill in the blanks:

1. ..... is simply a plan for a study.
2. ..... research is used to seek insights into general nature of the problem.

Notes

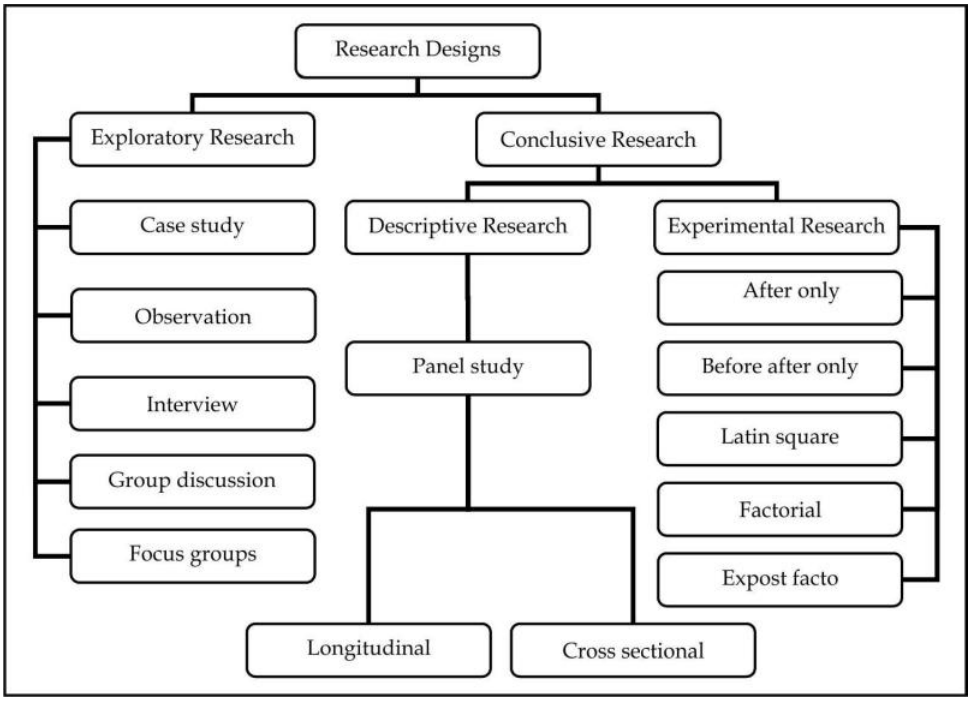
6.2 Types of Research Designs

However, a frequently used classification system is to group research designs under three major categories:

i. Research Designs in case of Exploratory Research Studies.

ii. Research Designs in case of Descriptive or Diagnostic Studies.

iii. Research Designs in case of Casual Research Studies.



6.3 Exploratory Research

The major emphasis in exploratory research is on converting broad, vague problem statements into small, precise sub-problem statements, which is done in order to formulate specific hypothesis. The hypothesis is a statement that specifies, "how two or more variables are related?"

In the early stages of research, we usually lack from sufficient understanding of the problem to formulate a specific hypothesis. Further, there are often several tentative explanations.

Examples:

1. "Sales are down because our prices are too high",
2. "our dealers or sales representatives are not doing a good job",
3. "our advertisement is weak" and so on.

In this scenario, very little information is available to point out, what is the actual cause of the problem. We can say that the major purpose of exploratory research is to identify the problem more specifically. Therefore, exploratory study is used in the initial stages of research.

Did u know? **Under what circumstances is exploratory study ideal?**

Notes

The following are the circumstances in which exploratory study would be ideally suited:

- To gain an insight into the problem
- To generate new product ideas
- To list all possibilities. Among the several possibilities, we need to prioritize the possibilities which seem likely
- To develop hypothesis occasionally
- Exploratory study is also used to increase the analyst's familiarity with the problem. This is particularly true, when the analyst is new to the problem area.

Example: A market researcher working for (new entrant) a company for the first time.

- To establish priorities so that further research can be conducted.
- Exploratory studies may be used to clarify concepts and help in formulating precise problems.

Example: The management is considering a change in the contract policy, which it hopes, will result in improved satisfaction for channel members. An exploratory study can be used to clarify the present state of channel members' satisfaction and to develop a method by which satisfaction level of channel members is measured

- To pre-test a draft questionnaire
- In general, exploratory research is appropriate to any problem about which very little is known. This research is the foundation for any future study.

Hypothesis Development at Exploratory Research Stage

At exploratory stage:

1. Sometimes, it may not be possible to develop any hypothesis at all, if the situation is being investigated for the first time. This is because no previous data is available.
2. Sometimes, some information may be available and it may be possible to formulate a tentative hypothesis.
3. In other cases, most of the data is available and it may be possible to provide answers to the problem.

Example: The example given below indicates each of the above type:

Research Purpose	Research Question	Hypothesis
(1) What product feature, if stated, will be most effective in the advertisement?	What benefit do people derive from this Ad appeal?	No hypothesis formulation is possible.
(2) What new packaging is to be developed by the company (with respect to a soft drink)?	What alternatives exist to provide a container for soft drink?	Paper cup is better than any other forms, such as a bottle.
(3) How can our insurance service be improved?	What is the nature of customer dissatisfaction?	Impersonalization is the problem.

Notes

In example 1, the research question is posed to determine "What benefit do people seek from the Ad?" Since no previous research is done on consumer benefit for this product, it is not possible to form any hypothesis.

In example 2, some information is currently available about packaging for a soft drink. Here it is possible to formulate a hypothesis which is purely tentative. The hypothesis formulated here may be only one of the several alternatives available.

In example 3, the root cause of customer dissatisfaction is known, i.e. lack of personalised service. In this case, it is possible to verify whether this is a cause or not.

6.3.1 Exploratory Research Methods

The quickest and the cheapest way to formulate a hypothesis in exploratory research is by using any of the four methods:

1. ***Literature Search:*** This refers to "referring to a literature to develop a new hypothesis". The literature referred are – trade journals, professional journals, market research finding publications, statistical publications etc.

Example: Suppose a problem is "Why are sales down?" This can quickly be analysed with the help of published data which should indicate "whether the problem is an "industry problem" or a "firm problem". Three possibilities exist to formulate the hypothesis.

1. The company's market share has declined but industry's figures are normal.
2. The industry is declining and hence the company's market share is also declining.
3. The industry's share is going up but the company's share is declining.

If we accept the situation that our company's sales are down despite the market showing an upward trend, then we need to analyse the marketing mix variables.

Example:

- A TV manufacturing company feels that its market share is declining whereas the overall television industry is doing very well.
- Due to a trade embargo imposed by a country, textiles exports are down and hence sales of a company making garment for exports is on the decline.

The above information may be used to pinpoint the reason for declining sales.

2. ***Experience Survey:*** In experience surveys, it is desirable to talk to persons who are well informed in the area being investigated. These people may be company executives or persons outside the organisation. Here, no questionnaire is required. The approach adopted in an experience survey should be highly unstructured, so that the respondent can give divergent views. Since the idea of using experience survey is to undertake problem formulation, and not conclusion, probability sample need not be used. Those who cannot speak freely should be excluded from the sample.

Example:

(1) A group of housewives may be approached for their choice for a "ready to cook product".

(2) A publisher might want to find out the reason for poor circulation of newspaper introduced recently. He might meet (a) Newspaper sellers (b) Public reading room (c) General public (d) Business community, etc.

These are experienced persons whose knowledge researcher can use.

Notes

3. ***Focus Group:*** Another widely used technique in exploratory research is the focus group. In a focus group, a small number of individuals are brought together to study and talk about some topic of interest. The discussion is co-ordinated by a moderator. The group usually is of 8-12 persons. While selecting these persons, care has to be taken to see that they should have a common background and have similar experiences in buying. This is required because there should not be a conflict among the group members on the common issues that are being discussed. During the discussion, future buying attitudes, present buying opinion etc., are gathered.

   Most of the companies conducting the focus groups first screen the candidates to determine who will compose the particular group. Firms also take care to avoid groups, in which some of the participants have their friends and relatives, because this leads to a biased discussion. Normallyl, a number of such groups are constituted and the final conclusions of various groups are taken for formulating the hypothesis. Therefore a key factor in focus group is to have similar groups. Normally there are 4-5 groups. Some of them may even have 6-8 groups. The guiding criterion is to see whether the latter groups are generating additional ideas or repeating the same with respect to the subject under study. When this shows a diminishing return from the group, the discussions stopped. The typical focus group lasts for 1-30 hours to 2 hours. The moderator under the focus group has a key role. His job is to guide the group to proceed in the right direction.

   The following should be the characteristics of a moderator/facilitator:

   - *Listening:* He must have a good listening ability. The moderator must not miss the participant's comment, due to lack of attention.
   - *Permissive:* The moderator must be permissive, yet alert to the signs that the group is disintegrating.
   - *Memory:* He must have a good memory. The moderator must be able to remember the comments of the participants. *Example:* A discussion is centered around a new advertisement by a telecom company. The participant may make a statement early and make another statement later, which is opposite to what was said earlier.

     For example: The participant may say that s(he) never subscribed to the views expressed in the advertisement by the competitor, but subsequently may say that the "current advertisement of competitor is excellent".
   - *Encouragement:* The moderator must encourage unresponsive members to participate.
   - *Learning:* He should be a quick learner.
   - *Sensitivity:* The moderator must be sensitive enough to guide the group discussion.
   - *Intelligence:* He must be a person whose intelligence is above the average.
   - *Kind/firm:* He must combine detachment with empathy.

Notes Variation of focus group

- ***Respondent moderator group:*** Under this method, the moderator will select one of the participants to act as a temporary moderator.
- ***Dualing moderator group:*** In this method, there are two moderators. They purposely take opposing positions on a given topic. This will help the researcher to obtain the views of both groups.

Contd...

Notes

- ***Two way focus group:*** Under this method one group will listen to the other group. Later, the second group will react to the views of the first group.
- ***Dual moderator group:*** Here, there are two moderators. One moderator will make sure that the discussion moves smoothly. The second moderator will ask a specific question.

4. ***Case studies:*** Analysing a selected case sometimes gives an insight into the problem which is being researched. Case histories of companies which have undergone a similar situation may be available. These case studies are well suited to carry out exploratory research. However, the result of investigation of case histories is always considered suggestive, rather than conclusive. In case of preference to "ready to eat food", many case histories may be available in the form of previous studies made by competitors. We must carefully examine the already published case studies with regard to other variables such as price, advertisement, changes in the taste etc.

A Case in Point

A company manufacturing electric shavers, known for its brand, wanted to introduce the product in Japan. Before the launch, the company made sure that all the 4Ps are acceptable to customers. When the product was launched, it met with failure. The company wondered what went wrong. Later investigations revealed that Japanese palms were very small and hence the product was not convenient for use. All possible causes were not listed and examined. This shows the importance of listing all factors during an exploratory research.

Self Assessment

Fill in the blanks:

3. The major emphasis in exploratory research is on converting ....., vague problem statements into ..... and ..... sub-problem statements.
4. In experience surveys, it is desirable to talk to persons who are well informed in the area being.....
5. Under the method of ..... group, the moderator will select one of the participants to act as a temporary moderator.
6. Most of the companies conducting the ..... groups first screen the candidates to determine who will compose the particular group.
7. The moderator must not miss the .....comment.
8. The moderator must encourage ..... members to participate.

6.4 Conclusive Research

Meaning: This is a research having clearly defined objectives. In this type of research, specific courses of action are taken to solve the problem.

In conclusive research, there are two types:

(a) Descriptive research

(b) Experimental research or Causal research.

Notes

6.4.1 Descriptive or Diagnostic Research

Meaning

(a) The name itself reveals that, it is essentially a research to describe something.

Example: It can describe the characteristics of a group such as—customers, organisations, markets etc.

Descriptive research provides "association between two variables" like income and place of shopping, age and preferences.

(b) Descriptive inform us about the proportions of high and low income customers in a particular territory. What descriptive research cannot indicate is that it cannot establish a cause and effect relationship between the characteristics of interest. This is the distinct disadvantage of descriptive research.

(c) Descriptive study requires a clear specification of *"Who, what, when, where, why and how"* of the research.

Example: Consider a situation of convenience stores (food world) planning to open a new outlet. The company wants to determine, "How people come to patronize a new outlet?"

Some of the questions that need to be answered before data collection for this descriptive study are as follows:

Who? Who is regarded as a shopper responsible for the success of the shop, whose demographic profile is required by the retailer?

What? What characteristics of the shopper should be measured?

Is it the age of the shopper, sex, income or residential address?

When? When shall we measure?

Should the measurement be made while the shopper is shopping or at a later time?

Where? Where shall we measure the shoppers?

Should it be outside the stores, soon after they visit or should we contact them at their residence?

Why? Why do you want to measure them?

What is the purpose of measurement? Based on the information, are there any strategies which will help the retailer to boost the sales? Does the retailer want to predict future sales based on the data obtained?

Answer to some of the above questions will help us in formulating the hypothesis.

How to measure? Is it a 'structured' questionnaire, 'disguised' or 'undisguised' questionnaire?

When to use descriptive study?

- To determine the characteristics of market such as:
  - (a) Size of the market
  - (b) Buying power of the consumer
  - (c) Product usage pattern
  - (d) To find out the market share for the product

Notes

- To determine the association of the two variables such as Ad and sales
- To make a prediction. We might be interested in sales forecasting for the next three years, so that we can plan for training of new sales representatives
- To estimate the proportion of people in a specific population, who behave in a particular way.

*Example:*What percentage of population in a particular geographical location would be shopping in a particular shop?

Management problem	Research problem	Hypothesis
How should a new product be distributed?	Where do customers buy a similar product right now?	Upper class buyers use 'Shopper's Stop' and middle class buyers buy from local departmental stores
What will be the target segment?	What kind of people buy our product now?	Senior citizens buy our products. Young and married buy our competitors products.

Task For each of the situation mentioned below, state whether the research should be exploratory, descriptive or causal:

1. To find out the relationship between promotion and sales.
2. To find out the consumer reaction regarding use of new detergents which are economical.
3. To identify the target market demographics, for a shopping mall.
4. Estimate the sales potential for ready-to-eat food in the northeastern parts of India.

Types of Descriptive Studies

There are two types of descriptive research:

(a) Longitudinal study

(b) Cross-sectional study

(a) ***Longitudinal Study:*** These are the studies in which an event or occurrence is measured again and again over a period of time. This is also known as 'Time Series Study'. Through longitudinal study, the researcher comes to know how the market changes over time.

Longitudinal studies involve panels. Panel once constituted will have certain elements. These elements may be individuals, stores, dealers etc. The panel or sample remains constant throughout the period. There may be some dropouts and additions. The sample members in the panel are being measured repeatedly. The periodicity of the study may be monthly or quarterly etc.

Example: Example for longitudinal study, assume a market research is conducted on ready to eat food at two different points of time T1 and T2 with a gap of 4 months. Each of the above two times, a sample of 2000 household is chosen and interviewed. The brands used most in the household is recorded as follows.

Notes

Brands	At T1	At T2
Brand X	500(25%)	600(30%)
Brand Y	700(35%)	650(32.5%)
Brand Z	400(20%)	300(15%)
Brand M	200(10%)	250(12.5%)
All others	200(10%)	250(12.5%)
	200	100%

As can be seen between period T1 and T2 Brand X and Brand M has shown an improvement in market share. Brand Y and Brand Z has decrease in market share, where as all other categories remains the same. This shows that Brand A and M has gained market share at the cost of Y and Z.

There are two types of panels:

- True panel
- Omnibus panel.

True panel: This involves repeat measurement of the same variables.

Example: Perception towards frozen peas or iced tea.

Each member of the panel is examined at a different time, to arrive at a conclusion on the above subject.

Omnibus panel: In omnibus panel too, a sample of elements is being selected and maintained, but the information collected from the member varies. At a certain point of time, the attitude of panel members "towards an advertisement" may be measured. At some other point of time the same panel member may be questioned about the "product performance".

Advantages of panel data:

- We can find out what proportion of those who bought our brand and those who did not. This is computed using the brand switching matrix.
- The study also helps to identify and target the group which needs promotional effort.
- Panel members are willing persons, hence a lot of data can be collected. This is because becoming a member of a panel is purely voluntary.
- The greatest advantage of panel data is that it is analytical in nature.
- Panel data is more accurate than cross-sectional data because it is free from the error associated with reporting past behaviour. Errors occur in past behaviour because of time that has elapsed or forgetfulness.

Disadvantages of panel data:

- The sample may not be representative. This is because sometimes, panels may be selected on account of convenience.
- The panel members who provide the data, may not be interested to continue as panel members. There could be dropouts, migration etc. Members who replace them may differ vastly from the original member.
- Remuneration given to panel members may not be attractive. Therefore, people may not like to be panel members.

Notes

- Sometimes the panel members may show disinterest and non-committed.
- A lengthy period of membership in the panel may cause respondents to start imagining themselves to be experts and professionals. They may start responding like experts and consultants and not like respondents. To avoid this, no one should be retained as a member for more than 6 months.

(b) ***Cross-sectional study:*** Cross-sectional study is one of the most important types of descriptive research, it can be done in two ways

- Field study
- Field survey

Field study: This includes a depth study. Field study involves an in-depth study of a problem, such as reaction of young men and women towards a product.

Example: Reaction of Indian men towards branded ready-to-wear suit. Field study is carried out in real world environment settings. Test marketing is an example of field study.

Field survey: Large samples are a feature of the study. The biggest limitations of this survey are cost and time. Also, if the respondent is cautious, then he might answer the questions in a different manner. Finally, field survey requires good knowledge like constructing a questionnaire, sampling techniques used, etc.

Example: Suppose the management believes that geographical factor is an important attribute in determining the consumption of a product, like sales of a woolen wear in a particular location. Suppose that the proposition to be examined is that, the urban population is more likely to use the product than the semi-urban population. This hypothesis can be examined in a cross-sectional study. Measurement can be taken from a representative sample of the population in both geographical locations with respect to the occupation and use of the products. In case of tabulation, researcher can count the number of cases that fall into each of the following classes:

- Urban population which uses the product - Category I
- Semi-urban population which uses the product - Category II
- Urban population which does not use the product - Category III
- Semi-urban population which does not use the product - Category IV

Here, we should know that the hypothesis need to be supported and tested by the sample data i.e., the proportion of urbanities using the product should exceed the semi-urban population using the product.

6.4.2 Survey

The survey is a research technique in which data are gathered by asking questions of respondents. Survey research is one of the most important areas of measurement in applied social research. The broad area of survey research encompasses any measurement procedures that involve asking questions of respondents. A "survey" can be anything form a short paper-and-pencil feedback form to an intensive one-on-one in-depth interview.

Types of Surveys

Surveys can be divided into two broad categories: the questionnaire and the interview. Questionnaires are usually paper-and-pencil instruments that the respondent completes.

Notes

Interviews are completed by the interviewer based on the respondent says. Sometimes, it's hard to tell the difference between a questionnaire and an interview. For instance, some people think that questionnaires always ask short closed-ended questions while interviews always ask broad open-ended ones. But you will see questionnaires with open-ended questions (although they do tend to be shorter than in interviews) and there will often be a series of closed-ended questions asked in an interview.

Survey research has changed dramatically in the last ten years. We have automated telephone surveys that use random dialing methods. There are computerized kiosks in public places that allows people to ask for input. A whole new variation of group interview has evolved as focus group methodology. Increasingly, survey research is tightly integrated with the delivery of service. Your hotel room has a survey on the desk. Your waiter presents a short customer satisfaction survey with your check. You get a call for an interview several days after your last call to a computer company for technical assistance. You're asked to complete a short survey when you visit a web site.

Selecting the Survey Method

Selecting the type of survey you are going to use is one of the most critical decisions in many social research contexts. You'll see that there are very few simple rules that will make the decision for you – you have to use your judgment to balance the advantages and disadvantages of different survey types.

Population Issues

He first set of considerations have to do with the population and its accessibility.

1. ***Can the population be enumerated?:*** For some populations, you have a complete listing of the units that will be sampled. For others, such a list is difficult or impossible to compile. For instance, there are complete listings of registered voters or person with active drivers licenses. But no one keeps a complete list of homeless people. If you are doing a study that requires input from homeless persons, you are very likely going to need to go and find the respondents personally. In such contexts, you can pretty much rule out the idea of mail surveys or telephone interviews.

2. ***Is the population literate?:*** Questionnaires require that your respondents can read. While this might seem initially like a reasonable assumption for many adult populations, we know from recent research that the instance of adult illiteracy is alarmingly high. And, even if your respondents can read to some degree, your questionnaire may contain difficult or technical vocabulary. Clearly, there are some populations that you would expect to be illiterate. Young children would not be good targets for questionnaires.

3. ***Are there language issues?:*** We live in a multilingual world. Virtually every society has members who speak other than the predominant language. Some countries (like Canada) are officially multilingual. And, our increasingly global economy requires us to do research that spans countries and language groups. Can you produce multiple versions of your questionnaire? For mail instruments, can you know in advance the language your respondent speaks, or do you send multiple translations of your instrument? Can you be confident that important connotations in your instrument are not culturally specific? Could some of the important nuances get lost in the process of translating your questions?

4. ***Will the population cooperate?:*** People who do research on immigration issues have a difficult methodological problem. They often need to speak with undocumented immigrants or people who may be able to identify others who are. Why would we expect those respondents to cooperate? Although the researcher may mean no harm, the

Notes

respondents are at considerable risk legally if information they divulge should get into the hand of the authorities. The same can be said for any target group that is engaging in illegal or unpopular activities.

5. ***What are the geographic restrictions?:*** Is your population of interest dispersed over too broad a geographic range for you to study feasibly with a personal interview? It may be possible for you to send a mail instrument to a nationwide sample. You may be able to conduct phone interviews with them. But it will almost certainly be less feasible to do research that requires interviewers to visit directly with respondents if they are widely dispersed.

Sampling Issues

The sample is the actual group you will have to contact in some way. There are several important sampling issues you need to consider when doing survey research.

1. ***What data is available? :*** What information do you have about your sample? Do you know their current addresses? Their current phone numbers? Are your contact lists up to date?

2. ***Can respondents be found?:*** Can your respondents be located? Some people are very busy. Some travel a lot. Some work the night shift. Even if you have an accurate phone or address, you may not be able to locate or make contact with your sample.

3. ***Who is the respondent?:*** Who is the respondent in your study? Let's say you draw a sample of households in a small city. A household is not a respondent. Do you want to interview a specific individual? Do you want to talk only to the "head of household" (and how is that person defined)? Are you willing to talk to any member of the household? Do you state that you will speak to the first adult member of the household who opens the door? What if that person is unwilling to be interviewed but someone else in the house is willing? How do you deal with multi-family households? Similar problems arise when you sample groups, agencies, or companies. Can you survey any member of the organization? Or, do you only want to speak to the Director of Human Resources? What if the person you would like to interview is unwilling or unable to participate? Do you use another member of the organization?

4. ***Can all members of population be sampled?:*** If you have an incomplete list of the population (i.e., sampling frame) you may not be able to sample every member of the population. Lists of various groups are extremely hard to keep up to date. People move or change their names. Even though they are on your sampling frame listing, you may not be able to get to them. And, it's possible they are not even on the list.

5. ***Are response rates likely to be a problem?:*** Even if you are able to solve all of the other population and sampling problems, you still have to deal with the issue of response rates. Some members of your sample will simply refuse to respond. Others have the best of intentions, but can't seem to find the time to send in your questionnaire by the due date. Still others misplace the instrument or forget about the appointment for an interview. Low response rates are among the most difficult of problems in survey research. They can ruin an otherwise well-designed survey effort.

Question Issues

Sometimes the nature of what you want to ask respondents will determine the type of survey you select.

Notes

1. ***What types of questions can be asked?:*** Are you going to be asking personal questions? Are you going to need to get lots of detail in the responses? Can you anticipate the most frequent or important types of responses and develop reasonable closed-ended questions?
2. ***How complex will the questions be?:*** Sometimes you are dealing with a complex subject or topic. The questions you want to ask are going to have multiple parts. You may need to branch to sub-questions.
3. ***Will screening questions be needed?:*** A screening question may be needed to determine whether the respondent is qualified to answer your question of interest. For instance, you wouldn't want to ask someone their opinions about a specific computer program without first "screening" them to find out whether they have any experience using the program. Sometimes you have to screen on several variables (e.g., age, gender, experience). The more complicated the screening, the less likely it is that you can rely on paper-and-pencil instruments without confusing the respondent.
4. ***Can question sequence be controlled?:*** Is your survey one where you can construct in advance a reasonable sequence of questions? Or, are you doing an initial exploratory study where you may need to ask lots of follow-up questions that you can't easily anticipate?
5. ***Will lengthy questions be asked?:*** If your subject matter is complicated, you may need to give the respondent some detailed background for a question. Can you reasonably expect your respondent to sit still long enough in a phone interview to ask your question?
6. ***Will long response scales be used?:*** If you are asking people about the different computer equipment they use, you may have to have a lengthy response list (CD-ROM drive, floppy drive, mouse, touch pad, modem, network connection, external speakers, etc.). Clearly, it may be difficult to ask about each of these in a short phone interview.

Content Issues

The content of your study can also pose challenges for the different survey types you might utilize.

1. ***Can the respondents be expected to know about the issue?:*** If the respondent does not keep up with the news (e.g., by reading the newspaper, watching television news, or talking with others), they may not even know about the news issue you want to ask them about. Or, if you want to do a study of family finances and you are talking to the spouse who doesn't pay the bills on a regular basis, they may not have the information to answer your questions.
2. ***Will respondent need to consult records?:*** Even if the respondent understands what you're asking about, you may need to allow them to consult their records in order to get an accurate answer. For instance, if you ask them how much money they spent on food in the past month, they may need to look up their personal check and credit card records. In this case, you don't want to be involved in an interview where they would have to go look things up while they keep you waiting (they wouldn't be comfortable with that).

Bias Issues

People come to the research endeavor with their own sets of biases and prejudices. Sometimes, these biases will be less of a problem with certain types of survey approaches.

1. ***Can social desirability be avoided?:*** Respondents generally want to "look good" in the eyes of others. None of us likes to look like we don't know an answer. We don't want to say anything that would be embarrassing. If you ask people about information that may

Notes

put them in this kind of position, they may not tell you the truth, or they may "spin" the response so that it makes them look better. This may be more of a problem in an interview situation where they are face-to face or on the phone with a live interviewer.

2. ***Can interviewer distortion and subversion be controlled?:*** Interviewers may distort an interview as well. They may not ask questions that make them uncomfortable. They may not listen carefully to respondents on topics for which they have strong opinions. They may make the judgment that they already know what the respondent would say to a question based on their prior responses, even though that may not be true.

3. ***Can false respondents be avoided?:*** With mail surveys it may be difficult to know who actually responded. Did the head of household complete the survey or someone else? Did the CEO actually give the responses or instead pass the task off to a subordinate? Is the person you're speaking with on the phone actually who they say they are? At least with personal interviews, you have a reasonable chance of knowing who you are speaking with. In mail surveys or phone interviews, this may not be the case.

Administrative Issues

Last, but certainly not least, you have to consider the feasibility of the survey method for your study.

1. ***Costs:*** Cost is often the major determining factor in selecting survey type. You might prefer to do personal interviews, but can't justify the high cost of training and paying for the interviewers. You may prefer to send out an extensive mailing but can't afford the postage to do so.

2. ***Facilities:*** Do you have the facilities (or access to them) to process and manage your study? In phone interviews, do you have well-equipped phone surveying facilities? For focus groups, do you have a comfortable and accessible room to host the group? Do you have the equipment needed to record and transcribe responses?

3. ***Time:*** Some types of surveys take longer than others. Do you need responses immediately (as in an overnight public opinion poll)? Have you budgeted enough time for your study to send out mail surveys and follow-up reminders, and to get the responses back by mail? Have you allowed for enough time to get enough personal interviews to justify that approach?

4. ***Personnel:*** Different types of surveys make different demands of personnel. Interviews require interviewers who are motivated and well-trained. Group administered surveys require people who are trained in group facilitation. Some studies may be in a technical area that requires some degree of expertise in the interviewer.

Clearly, there are lots of issues to consider when you are selecting which type of survey you wish to use in your study. And there is no clear and easy way to make this decision in many contexts. There may not be one approach which is clearly the best. You may have to make tradeoffs of advantages and disadvantages. There is judgment involved. Two expert researchers may, for the very same problem or issue, select entirely different survey methods. But, if you select a method that isn't appropriate or doesn't fit the context, you can doom a study before you even begin designing the instruments or questions themselves.

6.4.3 Observation Studies

An observational study draws inferences about the possible effect of a treatment on subjects, where the assignment of subjects into a treated group versus a control group is outside the

Notes

control of the investigator. This is in contrast with controlled experiments, such as randomized controlled trials, where each subject is randomly assigned to a treated group or a control group before the start of the treatment.

Observational studies are sometimes referred to as natural experiments or as quasi-experiments. These differences in terminology reflect certain differences in emphasis, but a shared theme is that the early stages of planning or designing an observational study attempt to reproduce, as nearly as possible, some of the strengths of an experiment.

Self Assessment

Fill in the blanks:

9. True panel involves .....measurement of the same variables.
10. Surveys can be divided into two broad categories: the ……and the interview.

Table 6.1: Differences between Exploratory Research and Descriptive Research

Exploratory research	Descriptive research
It is concerned with the "Why" aspect of consumer behaviour i.e., it tries to understand the problem and not measure the result.	It is concerned with the "What", "When" or "How often" on the consumer behaviour.
This research does not require large samples.	This needs large samples of respondents.
Sample need not be representing the population.	Sample must be representative of population.
Due to imprecise statement, data collection is not easy.	Statement is precise. Therefore data collection is easy.
Characteristics of interest to be measured is not clear.	Characteristics of interest to be measured is clear.
There is no need for a questionnaire for collecting the data.	There should be a properly designed questionnaire for data collection.
Data collection methods are: • Focus group • Literature Searching • Case study	• Use of panel data • Longitudinal • Cross-sectional studies

Task For the below mentioned scenario, lay down your recommendation of the most suitable type of research. Explain the reasons for your choice.

(1) Exploratory

(2) Descriptive

(3) Experimentation

(4) Longitudinal

(5) Cross-sectional

(a) A Tyre manufacturer is expecting recession in the next two years. The firm would like to know the changes that are to be made in the current marketing strategy, so as to minimize the adverse effect of the company's performance on account of recession.

Contd...

Notes

(b) A company manufacturing cell phones is concerned about a new brand being introduced by a competitor. The company would like to monitor how the new brand of the competitor will affect its market share in the next one year.

(c) A ready-to-eat food major would like to introduce iced tea. The company feels that this product is superior to what is already available in the market. The company wants to develop a unique promotional theme for the new product so that it may be clearly differentiated by the consumer and should appeal to broader section of the population.

(d) A co-operative bank has 4,000 customers who have taken personal loan or vehicle loan. Of late, the bank feels that there has been an increase in the number of defaulters. The bank would like to know whether people who are regular (no default) and defaulters differ in terms of characteristics such as age, income, occupation, sex, marital status.

6.5 Causal Research

Causal Research are the studies that engage in hypotheses testing usually explain the nature of certain relationships, or establish the differences among groups or the independence of two or more factors in a situation. A research design in which the major emphasis is on determining a cause-and-effect relationship. The research is used to measure what impact a specific change will have on existing norms and allows market researchers to predict hypothetical scenarios upon which a company can base its business plan.

Example: If a clothing company currently sells blue denim jeans, causal research can measure the impact of the company changing the product design to the colour white.

Following the research, company bosses will be able to decide whether changing the colour of the jeans to white would be profitable.

To summaries, causal research is a way of seeing how actions now will affect a business in the future. Nevertheless, it has to be remembered that not all causal research hypotheses can be studied. There are many reasons for this, one of them being that true random assignment is not possible in many cases. Gender cannot be randomly assigned, and therefore already you cannot test all causal hypotheses. The three main reasons why you can't test everything deal with

1. ***technology,*** or the impossibility by today's technology to be able to do certain tasks, such as assign gender.
2. ***ethics,*** because we can't randomly assign that some people receive a virus to test its effects, or that some participants have to act as slaves and others as masters to test a hypothesis, and
3. ***resources,*** if a researcher does not have the money or the equipment needed to perform a study, then it won't be done.

How to Prepare a Synopsis

Synopsis is an abstract form of research which underlines the research procedure followed and is presented before the guide for evaluating its potentiality. In one sentence it may be described as a condensation of the final report. The structure of synopsis varies and also depends on the guides' choice. However, for our understanding a common structure may be framed as under:

Notes

1. ***Defining the Problem:*** In defining the problem of the research objective, definition of key terms, general background information, limitations of the study and order of presentation should be mentioned in brief.

2. ***Review of Existing Literature:*** In this head, researcher should study the summary of different points of view on the subject matter as found in books, periodicals and approach to be followed at the time of writing.

3. ***Conceptual Framework and Methodology:*** Under this head the researcher should first make a statement of the hypothesis. Discussion on the research methodology used, duly pointing out the relationship between the hypothesis and objective of the study and finally discussions about the sources and means of obtaining data should also be made. In this head the researcher should also point out the limitations of methodology, if any, and the natural crises from which the research is bound to suffer for such obvious limitations.

4. ***Analysis of Data:*** Analysis of the data involves testing of hypothesis from data collected and key conclusions thus arrived.

5. ***General Conclusions:*** In general conclusions, the researcher should make a restatement of objectives. Conclusion with respect to the acceptance or rejection of hypothesis, conclusion with respect to the stated objectives, suggested areas of further research and final discussion of possible implications of the study for a model, group, theory and discipline.

Finally the researcher should mention about the bibliographies and appendices. The above format is drawn after a standard framework followed internationally in preparation of a synopsis. However, in our country, keeping in view the object of research, style and structure of synopsis varies and quite often it is found that the research guide exercises his own discretion in synopsis preparation than following some acceptable international norms. A standard format for preparation of synopsis commonly used in management and commerce research in India may be drawn as follows:

1. ***Introduction:*** This includes definition of the problem and its review from a historical perspective.

2. ***Objective of the Study:*** It defines the research purpose and its speciality from the existing available research in the related field.

3. ***Literature Review:*** It includes among other things, different sources from which the required abstract is drawn.

4. ***Methodology:*** It is intended to draw out the sequences followed in research and ways and manners of carrying out the survey and compilation of data.

5. ***Hypothesis:*** It is a formal statement relating to the research problem and it need to be tested based on the researchers' findings.

6. ***Model:*** It underlies the nature and structure of the model that the researcher is going to build in the light of survey findings.

Self Assessment

Fill in the blanks:

11. ……are the studies that engage in hypotheses testing usually explain the nature of certain relationships, or establish the differences among groups or the independence of two or more factors in a situation.

12. ……is an abstract form of research which underlines the research procedure followed and is presented before the guide for evaluating its potentiality.

Notes

6.6 Experimentation

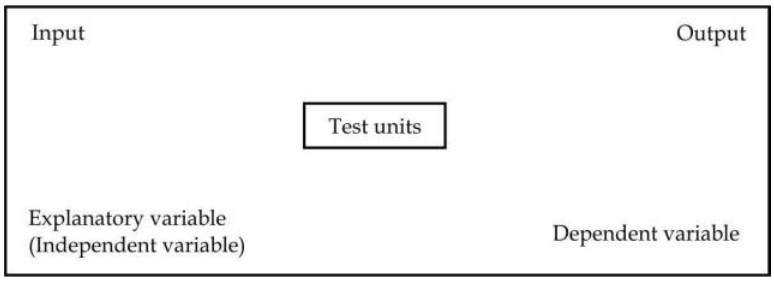
Experimentation Research is also known as causal research. Descriptive research, will suggest the relationship if any between the variable, but it will not establish cause and effect relationship between the variable.

Example: The data collected may show that the no. of people who own a car and their income has risen over a period of time. Despite this, we cannot say "No. of car increase is due to rise in the income". May be, improved road conditions or increase in number of banks offering car loans have caused in increase in the ownership of cars.

To find the causal relationship between the variables, the researcher has to do an experiment.

Example:

1. Which print advertisement is more effective? Is it front page, middle page or the last page?
2. Among several promotional measure, such as Advertisement, personal selling, "which one is more effective"? Can we increase sales of our product by obtaining additional shelf space? What is experimentation? It is research process in which one or more variables are manipulated, which shows the cause and effect relationship. Experimentation is done to find out the effect of one factor on the other. The different elements of experiment are explained below.



6.6.1 Test Units

These are units on which the experiment is carried out. It is done with one or more independent variables controlled by a person to find out its effect on a dependent variable.

6.6.2 Explanatory Variable

These are the variables whose effects the researcher wishes to examine. For example: Explanatory variables may be advertising, pricing, packaging etc.

6.6.3 Dependent Variable

This is a variable which is under study. For example: Sales, Consumer attitudes, Brand loyalty etc.

Example: Suppose a particular colour TV manufacturer reduces the price of the TV by 20%. Assume that his reduction is passed on to the consumer and expect the sales will go up by 15% in next one year. These types of experiments are done by leading TV companies during the festival season.

The causal research finds out whether the price reduction causes an increase in sales.

Notes

6.6.4 Extraneous Variables

These are also known as blocking variables. Extraneous variables affect the results of the experiments.

Example: Suppose a toffee manufacturing company is making an attempt to measure the response of the buyers to two different types of packaging, at two different locations. The manufacturer needs to keep other aspects the same, for each group of buyers. If the manufacturer allows the extraneous variable, namely, the price to vary between two buyer groups, then he will not be sure as to which particular packaging is preferred by the consumers.

Here the price change is an **extraneous** factor.

- There are two possible course of action with respect to extraneous variables.
- Extraneous variables may be physically controlled.

Example: Price in the above example.

In the second category, extraneous variables may totally elude the researcher's control. In this case, we say that the experiment has been confounded i.e., it is not possible to make any conclusions with regard to that experiment. Such a variable is known as "**Confounding variable**".

Example: A company introduces a product in two different cities. It would like to know the impact of advertising on sales. Simultaneously, the competitors' product in one of the cities is not available during this period due to a strike in the factory. Now, the researcher cannot conclude that sales of their product in that city have increased due to advertisement. Therefore, this experiment is confounded. In this case, the strike is the confounding variable.

Types of Extraneous Variables

The following are the various types of extraneous variables:

- ***History:*** History refers to those events which are external to the experiment, but occur at the same time as experiment is being conducted. This may affect the result.

Example: Let us suppose that, a manufacture makes a 20% cut in the price of a product and monitors sales in the coming weeks. The purpose of research is to learn about the impact of price on sales. Meanwhile, if the production of the product declines due to a shortage of raw materials, then the sales will not increase. Therefore, we cannot conclude that the price cut did not have any influence on sales because the **history** of external events have occurred during the period and we cannot control the event. The event can only be identified.

- ***Maturation:*** Maturation is similar to history. Maturation specifically refers to the changes occurring within the test units and not due to the effect of the experiment. Maturation takes place due to passage of time. It refers to the effect of people growing older. Persons who use a particular product may discontinue using that product and may switch over to an alternate product.

Example:

1. Pepsi is consumed when people are young. Due to passage of time, the consumer might prefer to consume Diet Pepsi or even avoid it altogether.

Notes

2. Assuming that a training programme is conducted for salesmen, the company wants to measure the impact of its sales programme. If the company finds that the sales have improved, it may not be due to its training programme. It may be because their salesmen have gained more experience now and know the customer better. Better understanding between salesmen and customer may be the reason for increased sales.

Maturation effect is not just limited to test unit, composed of people alone. Organisations also change, dealers grow, become more successful, diversify, and so on.

- ***Testing:*** Pre-testing effect occurs, when the same respondents are measured more than once. Responses given at a later stage will have a direct bearing on the responses given during an earlier measurement.

Example: Consider a respondent, who is given an initial questionnaire, intended to measure brand awareness. After examining him, if a second questionnaire similar to the initial questionnaire is given to the respondent, he will respond quite differently, because of the respondent's familiarity with the earlier questionnaire.

Pretest suffers from limitations of internal validity. This can be understood through an example. Assume that a respondent's opinion is measured before and after exposure to a TV commercial of Hyundai car with Shahrukh Khan as brand ambassador. When the respondent replies for second time, he may remember, how he rated Hyundai during the first measurement. He may give the same rating to simply prove that he is consistent. In that case, the difference between the two measurements will reveal nothing about the real impact.

Alternately, some of the respondents might give a different rating during the second measurement. This may not be due to the fact that the respondent has changed his opinion about Hyundai and the brand ambassador. He has given different rating because he does not want to be identified as a person with no change of opinion to the said commercial.

In both cases above, the internal validity suffers.

- ***Instrument Variation:*** Instrument variation effect is a threat to internal validity when human respondents are involved.

Example: An equipment such as vacuum cleaner is left behind for the customer's use for two weeks. After two weeks, respondents were given a questionnaire to answer. The reply may be quite different from what was before the trial of the product.

This may be because of two reasons:

(1) Some of the questions have been changed.

(2) The interviewers for pre-testing and post-testing periods are different.

The measurement in experiments will depend upon the instrument used for measurement. Also, results may vary due to the application of instruments, where there are several interviewers. Thus, it is very difficult to ensure that all the interviewers will ask the same questions with the same tone and develop the same rapport. There may be difference in response, because each interviewer conducts the interview differently.

- ***Bias in Selection:*** Bias in selection occurs because two groups selected for experiment may not be identical. If the two groups are asked various questions, they will respond differently. If multiple groups participate this error recurs frequently. There are two promotional advertisements, A and B, for 'ready to eat food'. The idea is to gauge the effectiveness of

Notes

two advertisements. Assume that the respondent exposed to 'A' are the dominant users of the product. Now, suppose 50% of those who saw Advertisement A bought the product and only 10% of those who saw Advertisement B bought the product. From the above, one should not conclude that advertisement 'A' is more effective than advertisement 'B'. The main difference may be due to food preference habits between the groups; even in this case, the internal validity might suffer but to a lesser degree.

- ***Experimental Mortality:*** Some members may leave the original group and some new members may join the old group. This is because some members might migrate to another geographical area. This change in composition of the members will alter the composition of the group itself.

Example: Assume that a vacuum cleaner manufacturer wants to introduce a new version. He interviews hundred respondents who are currently using the older version. Let us assume that, these 100 respondents have rated the existing vacuum cleaner on a 10 point scale (1 for lowest and 10 for highest). Let the mean rating of the respondents be 7.

Now the newer version is demonstrated to the same hundred respondents and the equipment is left with them for two months. At the end of two months, only 80 participants respond, since the remaining 20 refused to answer. Now the mean score of 80 respondents is 8 on the same 10 point scale. From this, can we conclude that the new vacuum cleaner is better?

The answer to the above question depends on the composition of 20 respondents who dropped out. Suppose the 20 respondents who dropped out displayed negative reaction to the product, then the mean score would not have been 8. It would have been even lower than 7. The difference in mean rating does not give the true picture. It does not indicate that the new product is better than the old one.

One might wonder, why not we leave the 20 respondents from the original group and calculate the mean rating of the remaining 80 and compare the two? But this method will also not solve the mortality effect. Mortality effect will occur in an experiment, irrespective of whether human beings are involved or not.

Task You are the manager of product planning and marketing research for a home appliance stores. Your company is considering a proposal to manufacture and market an emergency lamp in which segment the company currently does not have any product. You have assigned this project to one of your subordinates.

(a) Is this an exploratory, descriptive, or a causal study?

(b) What data would be useful for deciding whether to develop an emergency lamp or not?

(c) How will you design a study to obtain the needed data?

Self Assessment

Fill in the blanks:

13. ……Research is also known as causal research.
14. Extraneous Variables are also known as ……..variables.

Notes

15. Maturation is similar to …….
16. ……..in selection occurs because two groups selected for experiment may not be identical.

6.7 Experimental Designs

The various experimental designs are as follows:

- Purely post-design
- Before-after design
- Factorial design
- Latin square design
- Ex-post facto design

6.7.1 Purely Post-design

In this design, the dependent variable is measured after exposing the test units to the experimental variable. This can be understood with the help of following example:

Example: Assume M/s Hindustan Lever Ltd wants to conduct an experiment on the "Impact of free sample on the sale of toilet soaps". Small samples of toilet soaps are mailed to selected customers in a locality. After one month, a coupon of 25 paise off on one cake of soap is mailed to each customer to whom free samples were sent earlier. An equal number of these coupons are also mailed to people in another locality in the neighborhood. The coupons are coded to keep an account of the number of coupons redeemed from each locality. Suppose, 400 coupons were redeemed from the experimental group and 250 coupons were redeemed from the control group. The difference of 150 is supposed to be the effect of free samples. In this method, the conclusion can be drawn only after conducting the experiment.

6.7.2 Before-After Design

In this method, measurements are made before as well as after the design.

Example: Let us say that, an experiment is conducted to test an advertisement which is aimed at reducing alcoholism.

Attitudes and perceptions towards consuming liquor are measured before exposure to the advertisement. The group is exposed to an advertisement, which tells them the consequences, and their attitudes are again measured after several days. The difference, if any, shows the effectiveness of that advertisement.

The above example of "Before-after" suffers from validity threat due to the following.

Before Measure Effect

It alerts the respondents to the fact that they are being studied. The respondents may discuss the topics with friends and relatives and modify their behaviour accordingly.

Instrumentation Effect

This can be due to two different instruments being used – one before and one after. A change in the interviewers before and after, results in the instrumentation effect.

Notes

6.7.3 Factorial Design

Factorial design permits the researcher to test two or more variables at the same time. Factorial design helps to determine the effect of each of the variables and measure the interacting effect of many variables.

Example: A departmental store wants to study the impact of price reduction for a product. Given that, there is also promotion (POP) being carried out in the stores (a) near the entrance (b) at usual place, at the same time. Now assume that there are two price levels namely regular price A1 and reduced price A_2. Let there be three types of POP namely B_1, B_2 and B_3. There are 3 × 2 = 6 combinations possible. The combinations possible are B_1A_1, B_1A_2, B_2A_1, B_2A_2, B_3A_1, B3A2. Which of these combinations is best suited is what the researcher is interested in. Suppose there are 60 departmental stores of the chain divided into groups of 10 stores each. Now, randomly assign the above combination to each of these 10 stores as follows:

Combinations	Sales
B_1A_1	S_1
B_1A_2	S_2
B_2A_1	S_3
B_2A_2	S_4
B_3A_1	S_5
B_3A_2	S_6

S_1 to S_6 represents the sales resulting from each variable. The data gathered will provide details on product sales on account of two independent variables:

The two questions that will be answered are:

1. Is the reduced preice more effective than regular price?
2. Is the display at the entrance more effective than the display at the usual location? Also, the research will tell us about the interaction effect of the two variables.

Outcome of this experiment on sales is as follows:

1. Price reduction with display at the entrance.
2. Price reduction with display at the usual place.
3. No display and regular price applicable.
4. Display at the entrance with regular price applicable.

6.7.4 Latin Square Design

The researcher chooses three shelf arrangements in three stores. He would like to observe the sales generated in each of these stores at different periods. The researcher must make sure that one type of shelf arrangement is used in each store only once.

Notes

In the Latin Square design, only one variable is tested. As an example of Latin Square design, assume that a supermarket chain is interested in the effect of in-store promotion on sales. Suppose there are three promotions considered as follows:

1. No promotion.
2. Free sample with demonstration.
3. Window display.

Which of the three will be effective? The outcome may be affected by the size of the stores and the time period. If we choose three stores and three time periods, the total number of combination is 3 × 3 = 9. The arrangement is as follows:

Time period	Store		
	1	2	3
1	B	C	A
2	C	A	B
3	A	B	C

6.7.5 Ex-post Facto Design

This is a variation of "after only design". The groups such as experiment and control are identified only after they are exposed to the experiment.

Example: Let us assume that a magazine publisher wants to ascertain the impact of advertisement on knitting in *'Women's Era'* periodical. The subscribers were asked whether they have seen this advertisement on 'knitting'. Those who have read and not read were asked about the price, design etc. of the product. The difference indicates the effectiveness of the advertisement. In this design, the experimental group is set to receive the treatment rather than exposing it to the treatment by its choice.

Caselet

Different Opinion on Ad

A medium-size manufacturer of calculators was introducing a new scientific model. The company wants to communicate the same through an advertising programme. There was a discussion between the Marketing Manager and Vice-President — Marketing regarding this. The Marketing Manager was of the opinion that emphasis in the advertisement should be on features, since that would generate more sales. The Vice-President was of the opinion that the advertisement should emphasise on price, discounts etc. Since there was a difference in opinion, a market research agency was called and told to suggest a research design which would aid in making a final decision about the advertising programme.

If you were to head the Ad agency, what research designs would you recommend and why?

Self Assessment

Fill in the blanks:

17. In the Latin Square design, only ……..variable is tested.
18. ……..…Design is a variation of "after only design".

Notes

6.8 Summary

- There are primarily four types of research namely exploratory research, descriptive research and experimental research.
- Exploratory research helps the researcher to become familiar with the problem.
- It helps to establish the priorities for further research. It may or may not be possible to formulate Hypothesis during exploratory stage.
- To get an insight into the problem, literature search, experience surveys, focus groups, and selected case studies assist in gaining insight into the problem.
- The role of moderator or facilitator is extremely important in focus group. There are several variations in the formation of focus group.
- Descriptive research is rigid. This type of research is basically dependent on hypothesis.
- Descriptive research is used to describe the characteristics of the groups.
- It can also be used forecasting or prediction.
- True panel and Omni bus panel.
- In true panel same measurement are made during period of time.
- In Omni bus panel different measurement are made during a period of time.
- A cross-sectional study involves field study and field survey, the difference being the size of sample.
- Causal research is conducted mainly to prove the fact that one factor "X" the cause was responsible for the effect "Y".
- While conducting experiment, the researcher must guard against extraneous source of error.

6.9 Keywords

Causal Research: A research designed to determine cause and effect relationship.

Conclusive Research: This is a research having clearly defined objectives. In this type of research, specific courses of action are taken to solve the problem.

Descriptive Research: It is essentially a research to describe something.

Expost Facto Research: Study of the current state and factors causing it.

Extraneous Variable: These variables affect the response of test units. Also known as confounding variable.

Factorial Design: This is an experimental design when the effect of two or more variables are being studied simultaneously.

Field Study: Field study involves an in-depth study of a problem, such as reaction of young men and women towards a product.

Literature Research: It refers to "referring to a literature to develop a new hypothesis".

Longitudinal Study: These are the studies in which an event or occurrence is measured again and again over a period of time.

Notes

6.10 Review Questions

1. What are the different types of research design?
2. What is exploratory research? What methods are adopted during exploratory research?
3. What is descriptive research and what methods are adopted?
4. What is experimental research and what are the methods of conducting experimental research?
5. What are the types of errors that affect research design?
6. What is causal research?
7. How to solve the research problem systematically?
8. What are the different types of experimental designs?
9. Which type of research would you use to generate new product ideas and why?
10. Which type of research study would you use to determine the characteristics of market?

Answers: Self Assessment

1.	Research design	2.	Exploratory
3.	Broad, Small, Precise	4.	Investigated
5.	Respondent moderator	6.	Focus
7.	Participant's	8.	Unresponsive
9.	Repeat	10.	Questionnaire
11.	Causal Research	12.	Synopsis
13.	Experimentation	14.	Blocking
15.	History	16.	Bias
17.	One	18.	Ex-post Facto

6.11 Further Readings

Books

Cooper and Schinder, *Business Research Methods*, TMH.
CR Kotari, *Research Methodology*, Vishwa Prakashan.
David Luck and Ronald Rubin, *Marketing Research*, PHI.
Naresh Amphora, *Marketing Research*, Pearson Education.
S.N. Murthy & U. Bhojanna, *Business Research Methods*, 2nd Edition, Excel Books.
William MC Trochim, *Research Methods*, Biztantra.
William Zikmund, *Business Research Methods*, Thomson.

Online links

www.experiment-resources.com
www.socialresearchmethods.net

Unit 7: Sources and Methods of Data Collection

Notes

CONTENTS

Objectives
Introduction
7.1 Primary Data and Secondary Data
7.1.1 Primary Data
7.1.2 Secondary Data
7.2 Data Collection Methods
7.2.1 Observation Method
7.2.2 Qualitative Techniques of Data Collection
7.3 Questionnaire Designing
7.3.1 Importance of Questionnaire in MR
7.3.2 Developing a Good Questionnaire
7.3.3 Types of Questionnaires
7.3.4 Construction of Questionnaire Designing
7.3.5 Mail Questionnaire
7.3.6 Schedule Method
7.3.7 Sample Questionnaires
7.4 Summary
7.5 Keywords
7.6 Review Questions
7.7 Further Readings

Objectives

After studying this unit, you will be able to:

- Identify the types of data;
- Explain the data collection procedure for primary data and secondary data;
- Discuss the methods for data collection;
- Identify the meaning and process of questionnaire designing.

Introduction

Once the researcher has decided the 'Research Design', the next job is of data collection. For data to be useful, our observations need to be organized so that we can get some patterns and come to logical conclusions.

Statistical investigation requires systematic collection of data, so that all relevant groups are represented in the data.

Notes

To determine the potential market for a new product, for example, the researcher might study 500 consumers in a certain geographical area. It must be ascertained that the group contains people representing variables such as income level, race, education and neighborhood.

Caution The quality of data will greatly affect the conclusions and hence, utmost importance must be given to this process and every possible precaution should be taken to ensure accuracy, while gathering and collecting data.

Depending upon the sources utilized, whether the data has come from actual observations or from records that are kept for normal purposes, statistical data can be classified into two categories Primary and secondary.

7.1 Primary Data and Secondary Data

Data is one of the most important and vital aspect of any research studies. Researchers conducted in different fields of study can be different in methodology but every research is based on data which is analyzed and interpreted to get information. Data can be numbers, images, words, figures, facts or ideas. Data in itself cannot be understood and to get information from the data one must interpret it into meaningful information. There are various methods of interpreting data. Data sources are broadly classified into primary and secondary data. Let us discuss both of them:

7.1.1 Primary Data

The data directly collected by the researcher, with respect to the problem under study, is known as primary data. Primary data is also the firsthand data collected by the researcher for the immediate purpose of the study. Primary data is one which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by surveys conducted by individuals or research institutions.

Importance of Primary Data

Importance of Primary data cannot be neglected. A research can be conducted without secondary data but a research based on only secondary data is least reliable and may have biases because secondary data has already been manipulated by human beings. In statistical surveys it is necessary to get information from primary sources and work on primary data: for example, the statistical records of female population in a country cannot be based on newspaper, magazine and other printed sources. One such source is old and secondly they contain limited information as well as they can be misleading and biased.

1. ***Validity:*** Validity is one of the major concerns in a research. Validity is the quality of a research that makes it trustworthy and scientific. Validity is the use of scientific methods in research to make it logical and acceptable. Using primary data in research can improves the validity of research. First hand information obtained from a sample that is representative of the target population will yield data that will be valid for the entire target population.
2. ***Authenticity:*** Authenticity is the genuineness of the research. Authenticity can be at stake if the researcher invests personal biases or uses misleading information in the research. Primary research tools and data can become more authentic if the methods chosen to analyze and interpret data are valid and reasonably suitable for the data type. Primary sources are more authentic because the facts have not been overdone. Primary source can

Notes

be less authentic if the source hides information or alters facts due to some personal reasons. There are methods that can be employed to ensure factual yielding of data from the source.

3. ***Reliability:*** Reliability is the certainty that the research is enough true to be trusted on. For example, if a research study concludes that junk food consumption does not increase the risk of cancer and heart diseases. This conclusion should have to be drawn from a sample whose size, sampling technique and variability is not questionable. Reliability improves with using primary data. In the similar research mentioned above if the researcher uses experimental method and questionnaires the results will be highly reliable. On the other hand, if he relies on the data available in books and on internet he will collect information that does not represent the real facts.

Sources of Primary Data

Sources for primary data are limited and at times it becomes difficult to obtain data from primary source because of either scarcity of population or lack of cooperation. Regardless of any difficulty one can face in collecting primary data; it is the most authentic and reliable data source. Following are some of the sources of primary data:

1. ***Experiments:*** Experiments require an artificial or natural setting in which to perform logical study to collect data. Experiments are more suitable for medicine, psychological studies, nutrition and for other scientific studies. In experiments the experimenter has to keep control over the influence of any extraneous variable on the results.

2. ***Survey:*** Survey is most commonly used method in social sciences, management, marketing and psychology to some extent. Surveys can be conducted in different methods.

   - *Questionnaire:* It is the most commonly used method in survey. Questionnaires are a list of questions either open-ended or close-ended for which the respondents give answers. Questionnaire can be conducted via telephone, mail, live in a public area, or in an institute, through electronic mail or through fax and other methods.

   - *Interview:* Interview is a face-to-face conversation with the respondent. In interview the main problem arises when the respondent deliberately hides information otherwise it is an in depth source of information. The interviewer can not only record the statements the interviewee speaks but he can observe the body language, expressions and other reactions to the questions too. This enables the interviewer to draw conclusions easily.

   - *Observations:* Observation can be done while letting the observing person know that he is being observed or without letting him know. Observations can also be made in natural settings as well as in artificially created environment.

7.1.2 Secondary Data

Secondary data are statistics that already exist. They have been gathered not for immediate use. This may be described as "those data that have been compiled by some agency other than the user".

Did u know? **What is the categorization of Secondary data?**

Secondary data can be classified as:

- Internal secondary data
- External secondary data

Notes

Internal Secondary Data

Internal secondary data is a part of the company's record, for which research is already conducted. Internal data are those that are found within the organisation.

Example: Sales in units, credit outstanding, call reports of sales persons, daily production report, monthly collection report, etc.

External Secondary Data

The data collected by the researcher from outside the company. This can be divided into four parts:

- Census data
- Individual project report being published
- Data collected for sale on a commercial basis called syndicated data
- Miscellaneous data

Census data: Census data is the most important data among the sources of data. The following are some of the data that can obtained by census records:

- Census of the wholesale trade
- Census of the retail trade
- Population Census
- Census of manufacturing industries
- Individual project report being published
- Encyclopedia of business information sources
- Product finder
- Thomas registers etc.

Special Techniques of Market Research or Syndicated Data

These techniques involve data collection on a commercial basis i.e., data collected by this method is sold to interested clients, on payment. Example of such organisation is Neilson Retail, ORG Marg, IMRB etc. These organizations provide NRS called National Readership Survey to the sponsors and advertising agencies. They also provide business relationship survey called BRS which estimates the following:

(a) Rating

(b) Profile of the company etc.

(c) These people also provide TRP rating namely television rating points on a regular basis. This provides:

  (i) Viewership figures

  (ii) Duplication between programmes etc. Some of the interesting studies made by IMRB are SNAP – Study of Nations Attitude and Awareness Programme. In this study, the various groups of the Indian population and their life styles, attitudes of Indian housewives are detailed.

Notes

There is also a study called FSRP which covers children in the age group of 10-19 years. Beside their demographics and psychographics, the study covers those areas such as:

- Children as decision makers
- Role model of Indian children
- Pocket money and its usage
- Media reviews
- Favoured personalities and characteristics and
- Brand awareness and advertising recall

A syndicated source consists of market research firms offering syndicated services. These market research organisations, collects and updates information on a continues basis. Since data is syndicated, their cost is spread over a number of client organisations and hence cheaper. For example, a client firm can give certain specific question to be included in the questionnaire, which is used routinely to collect syndicated data. The client will have to pay extra for these. The data generated by these additional questions and analysis of such data will be revealed only to the firms submitting the questions. Therefore we can say, customization of secondary data is possible. Some areas of syndicated services are newspapers, magazine readership, TV channel popularity etc. Data from syndicated sources are available on a weekly or monthly basis.

Miscellaneous Secondary Data

Includes trade association such as FICCI, CEI, Institution of Engineers, chamber of Commerce, Libraries such as public library, University Library etc., literature, state and central government publications, private sources such as All India Management Association (AIMA), Financial Express and Financial Dailies, world bodies and international organizations such as IMF, ADB etc.

Advantages and Disadvantages of Secondary Data

Advantages

- It is economical, without the need to hire field staff.
- It saves time; (normally 2 to 3 months). If data is available on hand it can be tabulated in minutes.
- They provide information, which retailers may not be willing to reveal to researcher.
- No training is required to collect this data, unlike primary data.

Disadvantages

Because secondary data has been collected for some other projects, it may not fit in with the problem that is being defined. In some cases, the feed is so poor that the data becomes completely inappropriate. It may be ill-suited because of the following three reasons:

- ***Unit of Measurement:*** It is common for secondary data to be expressed in units.

Example: Size of the retail establishments, for instance, can be expressed in terms of gross sales, profits, square feet area and number of employees. Consumer incomes can be expressed in variables the individual, family, household etc. Secondary data available may not fit in easily.

Assume that the class intervals are quite different from those which are needed.

Notes

Example: Data available with respect to age group is as follows:

<18 year

18-24 years

25-34 years

35-44 years

Suppose the company needs a classification less than 20, 20-30 and 30-40, the above classification of secondary data cannot be used.

- ***Problem of Accuracy:*** The accuracy of secondary data available is highly questionable. A number of errors are possible in the collection and analysis of the data. Accuracy of secondary data depends upon:

  (a) ***Who has collected the data:*** The reliability of the source determines the accuracy of the data. Assume that a publisher of a private periodical conducts a survey of his readers. The main aim of the survey is to find out the opinion of readers about advertisements appearing in it. This survey is done by the publisher in the hope that other firms will buy this data before inserting advertisements.

  Assume that a professional M.R agency has conducted a similar survey and has sold its syndicated data on many periodicals.

  If you are an individual who wants information on a particular periodical you buy the data from M.R agency rather from the periodical's publisher. The reason for this is trust of the M.R agency. The reasons for trusting the M.R agency are as follows:

  1. Being an independent agency there is no bias. The M.R agency is likely to provide an unbiased data.
  2. The data quality of MR agency will be good since they are professionals.

  (b) ***How the data collected ?***

  1. What instruments were used?
  2. What type of sampling was done?
  3. How large was the sample?
  4. What was the time period of data collection? For example, days of the week, time of the day.

- ***Recency:*** This pertains to "how old was the information?" If it is five years old, it may be useless. Therefore, the publication lag is a problem.

Methods for Secondary Data Collection

The sources of unpublished data are many; they may be found in diaries, letters, unpublished biographies and autobiographies and may also be available with scholars and research workers, trade associations, labour bureaus and other public/private individuals and organizations.

Before using secondary data, the researcher must ensure the reliability, suitability and adequacy of data.

Notes

Secondary Data – Internal

Internal records or published records are often capable of giving remarkably useful information. Sometimes, the information may be sufficient enough to give the desired result. However, this preliminary information shall most of the time help in developing the overall research strategy and hence must be undertaken before any further research is contemplated.

For a manufacturing industry, for example, the internal production and sales records, if designed and maintained properly, can help in a big way even for formulating the companies strategies.

Secondary Data – External

External sources of data include statistics and reports issued by governments, trade associations and other reputable organizations such as advertising agencies and research companies and trade directories.

Did u know? **What are the sources of secondary data in India?**

In India, some of the major sources of secondary data are:

Indian Council of Agriculture, Central Statistical Organization, Army Statistical Organizations, National Accounts Statistics, Bulletin on Food Statistics, Handbook of Statistics on Small Scale Industries, RBI Bulletin, Annual survey of industries, Indian Labour Year Book, etc.

Task List some major secondary sources of information for the following:

1. Market research manager of a tea manufacturing company has to prepare a comprehensive report on the tea industry as a whole.
2. M.T.R has several product ideas on ready-to-eat products. It wishes to convert ideas into products and enter the market. Before entering, the company needs to find necessary information to assess the market potential.
3. An MNC wishes to open a showroom in a Metro. The first step that the company would like to take is to collect the information about suitability.
4. Number of residential houses less than 10 years old in a given locality.
5. Number of consultancy/recruitment firms in a city.
6. Percentage of families with children less than 15 years in a given locality.
7. Citizens who have electoral I.D cards in a local city.
8. Annual sales figures of a multi-retail outlet.

Self Assessment

Fill in the blanks:

1. ..... is also the first hand data collected by the researcher for the immediate purpose of the study.
2. ……can be at stake if the researcher invests personal biases or uses misleading information in the research.

Notes

3. ……is a part of the company's record, for which research is already conducted.

4. In ……the main problem arises when the respondent deliberately hides information otherwise it is an in depth source of information.

7.2 Data Collection Methods

Observation and questioning are two broad approaches available for primary data collection. The major difference between the two approaches is that, in questioning process, respondent play an active role, because of interaction with the researcher.

7.2.1 Observation Method

In observation method, only present/current behaviour can be studied. Therefore many researchers feel that this is a great disadvantage. A causal observation can enlighten the researcher to identify the problem. Such as length of the queue in front of a food chain, price and advertising activity of the competitor etc. observation is the least expensive of data collection.

Example 1: Suppose a safety week is celebrated and public is made aware of safety precautions to be observed while walking on the road. After one week, an observer can stand at a street corner and observe the No. of people walking on footpath and those walking on the road during a time period.

This will tell him whether the campaign on safety is successful or unsuccessful. Sometimes observation will be the only method available to the researcher.

Example 2: Behaviour or attitude of children, and also of those who are inarticulate.

Types of Observation Methods

There are several methods of observation of which, any one or a combination of some of them, can be used by the observer. They are:

- Structured or unstructured observation methods
- Disguised or undisguised observation methods
- Direct-indirect observation
- Human-mechanical observation

1. ***Structured-unstructured Observation Methods:*** Whether the observation should be structured or unstructured depends on the data needed.

Example 1: A Manager of a hotel wants to know "How many of his customers visit the hotel with family and how many visits as single customer".

Here observation is structured, since it is clear "what is to be observed". He may tell the waiters to record this. This information is required to decide the tables and chairs requirement and also the layout.

Suppose, the Manager wants to know how single customer and customer with family behave and what is their mood. This study is vague, it needs non-structured observation.

It is easier to record structured observation than non structured observation.

Notes

Example 2: To distinguish between structured and unstructured observation, consider a study, investigating the amount of search that goes into a "soap purchase". On the one hand, the observers could be instructed to stand at one end of a supermarket and record each sample customer's search. This may be observed and recorded as follows. "Purchaser first paused after looking at HLL brand". He looked at the price on of the product, kept the product back on the shelf, then picked up a soap cake of HLL and glanced at the picture on the pack and its list of ingredients, and kept it back. He then checked the label and price for P&G product, kept that back down again, and after a slight pause, picked up a different flavor soap of M/S Godrej company and placed it in his trolley and moved down the aisle. On the other hand, observers might simply be told to record the "First soap cake examined", by checking the appropriate boxes in the observation form. The "second situation" represents more structured than the first.

To use more structured approach, it would be necessary to decide precisely, what is to be observed and the specific categories and units that would be used to record the observations.

2. ***Disguised-undisguised Observation Methods:*** In disguised observation, the respondents do not know that they are being observed. In non disguised observation, the respondents are well aware that they are being observed. In disguised observation, many times observers pose as shoppers. They are called as "mystery shoppers". They are paid by the research organisations. The main strength of disguised observation is that, it allows for maintaining the true reactions of the individuals.

   In undisguised method, observation may be contained due to induced error by the objects of observation. The ethical aspect of disguised observations is still questionable.

3. ***Direct-indirect Observation:*** In direct observation, the actual behaviour or phenomenon of interest is observed. In Indirect observation, results of the consequences of the phenomenon are observed. Suppose, researcher is interested in knowing about the soft drink consumption of a student in a hostel room. He may like to observe empty soft drink bottles dropped into the bin. Similarly, the observer may seek the permission of the hotel owner, to visit the kitchen or stores. He may carry out a kitchen/stores audit, to find out the consumption of various brands of spice items being used by the Hotel.

   Notes It may be noted that, the success of an indirect observation largely depends on "How best the observer is able to identify physical evidence of the problem under study".

4. ***Human-mechanical Observation:*** Most of the studies in marketing research based on human observation, wherein trained observers are required to observe and record their observations. In some cases, mechanical devices such as eye cameras are used for observation. One of the major advantages of electrical/mechanical devices is that, their recordings are free from subjective bias.

Advantages of Observation Method

1. The original data can be collected at the time of occurrence of the event.
2. Observation is done in natural surroundings. Therefore facts are known, where questionnaire, experiments have environmental as well as time constraint.
3. Sometimes the respondents may not like to part with some of the information. Those information can be got by the researcher by observation.
4. Observation can be done on those who cannot articulate.
5. Bias of the researcher is greatly reduced in observation method.

Notes

Limitations of Observation Method

1. The observer might be waiting at the point of observation. Still the desired event may not take place i.e. observation is required over a long period of time and hence delay may occur.
2. For observation, extensive training of observers is required.
3. This is an expensive method.
4. External observation gives only surface indications. To go beneath the surface it is very difficult. So only overt behaviour can be observed.
5. Two observers may observe the same event but may draw inference differently.
6. It is very difficult to gather information on (1) Opinions (2) Intentions etc.

Task What observation technique would you use to gather the following information?

1. What kind of influence do children have on the purchase behaviour of their parents?
2. How do discounts influence the purchase behaviour of customers buying colour TV?
3. A study to find out the potential location for a snack bar in a city.

7.2.2 Qualitative Techniques of Data Collection

Qualitative research is used to analyse those data which cannot be quantified. Qualitative research is used in exploratory research. The number of respondents covered in this type of research is small compared to quantitative research.

There are four major techniques in Qualitative Research. They are:

1. Depth Interview
2. Delphi Technique
3. Focus Group
4. Projective Technique

Depth Interview

Unstructured, direct interview is known as a depth interview. Here the interviewer will continue to ask probing questions of like, "What did you mean by that statement?", "Why did you feel this way?" and "What other reasons do you have?" etc., until he is satisfied that he has obtained the information he wants.

Notes The unstructured interview is free from restrictions imposed by a formal list of questions.

The interview may be conducted in a casual and informal manner in which the flow of the conversation determines what questions are to be asked and the order in which they should be asked.

Notes

Delphi Technique

This is a process where a group of experts in the field gather together. They may have to reach a consensus on forecasts. Sometimes, the judgment may be made by some group members who have strong personalities.

Notes In the Delphi approach, the group members are asked to make individual judgments about a particular subject, say 'sales forecast'.

These judgments are compiled and returned to the group members, so that they can compare their previous judgment with those of others. Then they are given an opportunity to revise their judgments, especially if it differs from the others. They can say, why their judgment is accurate, even if it differs, from that of the other group members. After 5 to 6 rounds of interaction, the group members reach conclusion.

Focus Group Interview

They are the best known and most widely used type of indirect interviews. Here, a group of people jointly participate in an unstructured indirect interview conducted by a moderator. The group usually consists of six to ten people. In general, the selected persons have similar backgrounds. The moderator attempts to focus the discussion on the problem areas.

Focus groups are used primarily to provide background information and to generate hypothesis rather than to provide solution to problems. The areas of application include:

(1) Development of new product concept.

(2) The generation of ideas for improving established products.

(3) Development of creative concepts in advertising.

An example of the use of the focus group technique in the development of advertising may be looked at. Assume that company X wants to introduce electrical cars. Just prior to the introduction of the new car, the company conducts two focus group interviews to see "what is the dealers' perception about key benefits of the new type of car?" Assume that previous research indicated the customers would buy the new car, provided they were less expensive than the conventional cars. Since the new car was priced lower than price of a conventional car, the company expected no problems with the dealers accepting the new car. Instead, the focus group interviews found that the dealers were doubtful about the acceptance of electrical car in the Indian market, since it is new, despite the fact that it is cheaper than regular car. Customers were concerned about charging mode, facilities for doing so, battery life and above all, newness of the concept.

Projective Techniques

Projective techniques (Indirect method of gathering information/indirect interview) are unstructured and involve indirect form of questioning.

In projective techniques, respondents are asked to interpret the behaviour of users, rather than describe their own behaviour. In interpreting the behaviour of others, respondents indirectly project their own motivation and feelings into the situation.

Example: Many a time, people do not want to reveal their true motive for fear of being branded 'old fashioned'. Questions such as "Do you do all household work yourself?" The

Notes

answer may be 'no', though the truth is 'yes'. A 'yes' answer may not be given because it may suggest that the family is not financially sound and cannot afford a maid for help.

Two types of projective techniques are available:

1. An ambiguous stimulus is presented to respondents.
2. In reacting to the stimulus, the respondents will indirectly reveal their own feeling.

The general categories of projective techniques are:

1. Word association test
2. Completion technique
3. TAT and
4. Cartoon test

1. ***Word Association Test:*** This test consists of presenting a series of stimulus words to the respondent, who is asked to answer quickly with the first word that comes to his mind. The respondent, by answering quickly, gives the word that s(he) or she associates most closely with the stimulus word.

Example:

- What brand of detergent comes to your mind first, when I mention washing of an expensive cloth?
  - (1) Surf
  - (2) Tide
  - (3) Key
  - (4) Ariel
- Who drinks the milk most?
  - (1) Athletes
  - (2) Young boys
  - (3) Adults
  - (4) Children
- In a study of cigarettes, the respondent is asked to give the first word that comes to his mind.
  - (1) Injurious
  - (2) Style
  - (3) Strong
  - (4) Stimulus
  - (5) Bad manners
  - (6) Disease
  - (7) Pleasure

2. ***Completion Techniques***
   - Sentence completion
   - Story completion

Notes

- *Sentence Completion:* Here the respondents have to finish a set of incomplete sentences. *Example:* Let us make a study dealing with people's inner feelings towards software professionals.

  (a) Earnings of a software professional …….

  (b) Being a software professional means …….

  (c) Working hours for software professional are …….

  (d) The personal life of a software professional is …….

  (e) The social status of software professional is …….

  Suppose you want to study the attitude towards a periodical:

  (a) A person who reads *Women's Era periodical* is …….

  (b) *Business World periodical* appeals to ……..

  (c) *Outlook periodical* is read by …….

  (d) *Investor periodical* is mostly liked by …….

  Suppose you want to provide a basis for developing advertising appeal for a brand of cooking oil, the following sentence may be used:

  (a) People use cooking oil …….

  (b) Most of the new cooking oil ……

  (c) Costliest cooking oil ……

  (d) The thing I enjoy about cooking oil used by my family……..

  (e) One important feature to be highlighted in the advertisement about cooking oil is…….

- *Story Completion:* A situation is described to a respondent who is asked to complete the story based on his opinion and attitude. This technique will reveal the interest of the respondent, but it is difficult to interpret.

Example:

- Mr. X belongs to the upper-middle class. He received a telephone call, where the caller said that "I am from Globe Travels. Sir, I want to tell you about our recent offer, that is, if you travel to the US this summer, you will get two tickets free by the year end to fly to the Far East.

  What was Mr. X's reaction? Why?

- Two children are quarreling at the breakfast table before going to school. The younger of the two, has spilled coffee on her brother's shirt which he was supposed to wear on the same day for attending annual sports event.

  What will the mother do?

The story completion has numerous applications in solving marketing problem. The most important of which is to provide data to the seller, recognising the image and feelings people have about the company's products and services. This method is used before finalising an advertisement.

3. ***Thematic Apperception Test (TAT):*** TAT is a projective technique. It is used to measure the attitude and perception of the individual. Some picture cards are shown to respondents.

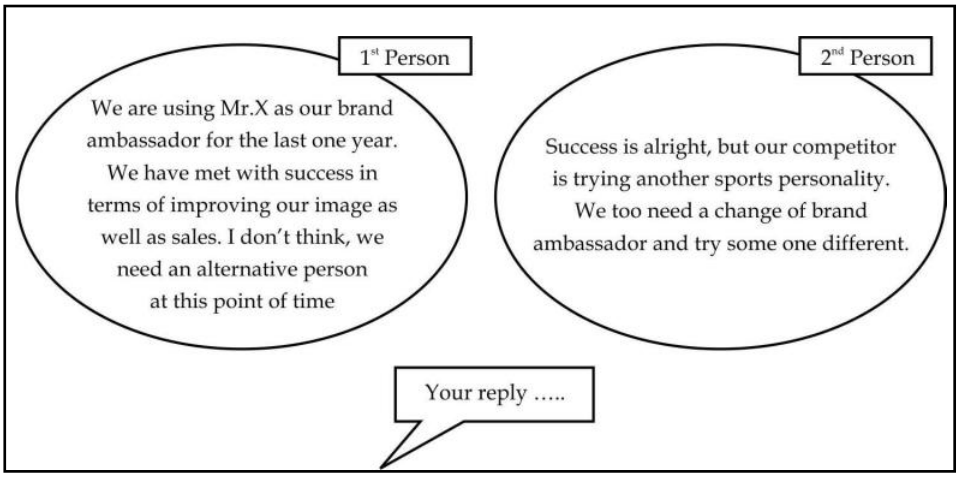
Notes

The respondent is required to tell the story by looking at the picture. When the subjects start telling the story, the researcher notices the respondents' expression, pauses and emotions to draw the inference.

In the TAT, the test subject (the boy shown here) examines a set of cards that portray human figures in a variety of settings and situations, and is asked to tell a story about each card. The story includes the event shown in the picture, preceding events, emotions and thoughts of those portrayed, and the outcome of the event shown. The story content and structure are thought to reveal the subject's attitudes, inner conflicts, and views. Customer insights may be extracted by posing the questions given above to the respondents

Source: (http://www.minddisorders.com)

4. ***Cartoon Test or Balloon Test:*** Here a cartoon is shown. The cartoon character belongs to a particular situation. One or more of 'balloons' include the conversation of the character, and is left open and the respondent is asked to fill in. In comparing the cartoon technique with the direct question, let us take the example of "choosing a brand ambassador".



In the above case, with which person would you agree and why ?

Caselet

Given below are some topics. In each case, indicate whether the research is qualitative or quantitative in nature. Also recommend specific techniques for each.

(a) A company would like to come out with ideas to creatively communicate the benefits of a new detergent through a TV commercial.

(b) Hospital authorities want to ascertain their patients' ratings of attributes like medical treatment, room service, emergency service, etc.

Contd...

Notes

(c) After discussing with several sales people, the sales manager suspects that the morale of the sales force is low, and wants to confirm this by using an employee morale questionnaire.

(d) A firm marketing toffee has two alternative wrapper designs for the product and is wonders, which one will result in higher sales.

Self Assessment

Fill in the blanks:

5. ……and ……..are two broad approaches available for primary data collection.
6. In……, the actual behaviour or phenomenon of interest is observed.
7. …….is used to analyse those data which cannot be quantified.
8. Unstructured, direct interview is known as a …….interview.

7.3 Questionnaire Designing

Questionnaires are an inexpensive way to gather data from a potentially large number of respondents. Often they are the only feasible way to reach a number of reviewers large enough to allow statistically analysis of the results. A well-designed questionnaire that is used effectively can gather information on both the overall performance of the test system as well as information on specific components of the system. If the questionnaire includes demographic questions on the participants, they can be used to correlate performance and satisfaction with the test system among different groups of users. A questionnaire is a research instrument consisting of a series of questions and other prompts for the purpose of gathering information from respondents. Although they are often designed for statistical analysis of the responses, this is not always the case. The questionnaire was invented by Sir Francis Galton.

Caution It is important to remember that a questionnaire should be viewed as a multi-stage process beginning with definition of the aspects to be examined and ending with interpretation of the results.

Every step needs to be designed carefully because the final results are only as good as the weakest link in the questionnaire process. Although questionnaires may be cheap to administer compared to other data collection methods, they are every bit as expensive in terms of design time and interpretation.

Questionnaires have advantages over some other types of surveys in that they are cheap, do not require as much effort from the questioner as verbal or telephone surveys, and often have standardized answers that make it simple to compile data. However, such standardized answers may frustrate users. Questionnaires are also sharply limited by the fact that respondents must be able to read the questions and respond to them. Thus, for some demographic groups conducting a survey by questionnaire may not be practical.

Notes As a type of survey, questionnaires also have many of the same problems relating to question construction and wording that exist in other types of opinion polls.

Notes

7.3.1 Importance of Questionnaire in MR

To study:

1. Behaviour, past and present.
2. Demographic characteristics such as age, sex, income, occupation.
3. Attitudes and opinions.
4. Level of knowledge.

7.3.2 Developing a Good Questionnaire

1. It must be simple. The respondents should be able to understand the questions.
2. It must generate replies that can be easily be recorded by the interviewer.
3. It should be specific, so as to allow the interviewer to keep the interview to the point.
4. It should be well arranged, to facilitate analysis and interpretation.
5. It must keep the respondent interested throughout.

7.3.3 Types of Questionnaires

1. Structured and non-disguised
2. Structured and disguised
3. Non-structured and Disguised
4. Non-structured and Non-disguised

1. ***Structured and Non-disguised Questionnaire:*** Here, questions are structured so as to obtain the facts. The interviewer will ask the questions strictly in accordance with the pre-arranged order.

Example: What are the strengths of soap A in comparison with soap B?

- Cost is less
- Lasts longer
- Better fragrance
- Produces more lather
- Available in more convenient sizes

Structured and non-disguised questionnaire is widely used in market research. Questions are presented with exactly the same wording and same order to all respondents. The reason for standardizing the question is to ensure that all respondents reply the same question. The purpose of the question is clear. The researcher wants the respondent to choose one of the five options given above. This type of questionnaire is easy to administer. The respondents have no difficulty in answering, because it is structured, the frame of reference is obvious.

In a non-disguised type, the purpose of the questionnaire is known to the respondent.

Notes

Example: "Subjects attitude towards Cyber laws and the need for government legislation to regulate it".

- Certainly, not needed at present
- Certainly not needed
- I can't say
- Very urgently needed
- Not urgently needed

2. ***Structured and disguised Questionnaire:*** This type of questionnaire is least used in marketing research. This type of questionnaire is used to know the peoples' attitude, when a direct undisguised question produces a bias. In this type of questionnaire, what comes out is "what does the respondent know" rather than what he feels. Therefore, the endeavour in this method is to know the respondent's attitude.

   Currently, the "Office of Profit" Bill is:

   (a) In the Lok Sabha for approval.

   (b) Approved by the Lok Sabha and pending in the Rajya Sabha.

   (c) Passed by both the Houses, pending the presidential approval.

   (d) The bill is being passed by the President.

   Depending on which answer the respondent chooses, his knowledge on the subject is classified.

   In a disguised type, the respondent is not informed of the purpose of the questionnaire. Here the purpose is to hide "what is expected from the respondent?"

Example:

1. "Tell me your opinion about Mr. Ben's healing effect show conducted at Bangalore?"
2. "What do you think about the Babri Masjid demolition?"

3. ***Non-structured and Disguised Questionnaire:*** The main objective is to conceal the topic of enquiry by using a disguised stimulus. Though the stimulus is standardized by the researcher, the respondent is allowed to answer in an unstructured manner. The assumption made here is that individual's reaction is an indication of respondent's basic perception. Projective techniques are examples of non-structured disguised technique. The techniques involve the use of a vague stimulus, which an individual is asked to expand or describe or build a story, three common types under this category are (a) Word association (b) Sentence completion (c) Story telling.

4. ***Non-structured and Non-disguised Questionnaire:*** Here the purpose of the study is clear, but the responses to the question are open-ended.

Example: "How do you feel about the Cyber law currently in practice and its need for further modification"? The initial part of the question is consistent. After presenting the initial question, the interview becomes very unstructured as the interviewer probes more deeply. Subsequent answers by the respondents determine the direction the interviewer takes next. The question asked by the interviewer varies from person to person. This method is called "the depth interview". The major advantage of this method is the freedom permitted to the interviewer. By not restricting the respondents to a set of replies, the experienced interviewers will be above to get the information from the respondent fairly and accurately.

Notes

Caution The main disadvantage of this method of interviewing is that it takes time, and the respondents may not co-operate.

Another disadvantage is that coding of open-ended questions may pose a challenge.

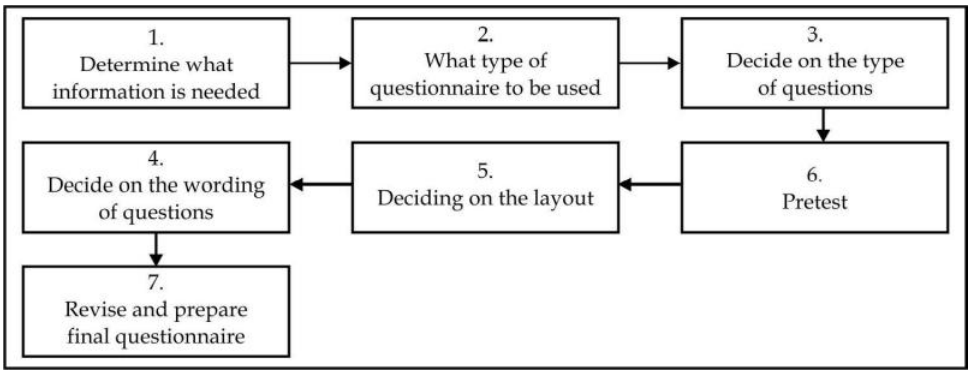
Example: When a researcher asks the respondent "Tell me something about your experience in this hospital". The answer may be "Well, the nurses are slow to attend and the doctor is rude. 'Slow' and 'rude' are different qualities needing separate coding. This type of interviewing is extremely helpful in exploratory studies.

Table 7.1: Types of Questionnaires

Types	Characteristics
Structured – Disguised	• The same question is posed to each respondent. • Administering the questionnaire and post-administration work is simple i.e. coding tabulating, etc. is easy. • This type of questionnaire is least used in market research. • Respondents' bias is minimized.
Unstructured – Disguised	• This type of questionnaire is very commonly used for focus group discussions. • This is difficult to analyse, code etc, • No fixed set of questions. • The inner self (why) of an individual is researched. Eg: Motivation Research.
Unstructured – Undisguised	• No fixed questions. • Suitable for conducting depth interview. • Subject-matter can be questioned in great detail. • Coding, tabulating etc. are difficult not a very frequently used method.
Structured – Undisguised	• Fixed set of questions to every respondent. • Inappropriate when researcher wants to probe deeper. • Easy to administer coding, tabulating is easy. • Due to structuring and undisguised nature of the questionnaire, there is no possibility of the respondent misunderstanding the question. This is the most commonly used method.

7.3.4 Construction of Questionnaire Designing

The following are the seven steps:



Notes

Determine What Information is required

The first question to be asked by the market researcher is "what type of information does he need from the survey?" This is valid because if he omits some information on relevant and vital aspects, his research is not likely to be successful. On the other hand, if he collects information which is not relevant, he is wasting his time and money.

Caution At this stage, information required, and the scope of research should be clear.

Therefore the steps to be followed at the planning stage are:

1. Decide on the topic for research.
2. Get additional information on the research issue, from secondary data and exploratory research. The exploratory research will suggest "what are the relevant variables?"
3. Gather what has been the experience with similar study.
4. The type of information required. There are several types of information such as (a) awareness, (b) facts, (c) opinions, (d) attitudes, (e) future plans, (f) reasons.

Facts are usually sought out in marketing research.

Example: Which television programme did you see last Saturday? This requires a reasonably good memory and the respondent may not remember. This is known as recall loss. Therefore questioning the distant past should be avoided. Memory of events depends on (1) Importance of the events and (2) Whether it is necessary for the respondent to remember. In the above case, both the factors are not fulfilled. Therefore, the respondent does not remember. On the contrary, a birthday or wedding anniversary of individuals is remembered without effort since the event is important. Therefore, the researcher should be careful while asking questions about the past. First, he must make sure that the respondent has the answer.

Example: Do you go to the club? He may answer 'yes', though it is untrue. This may be because the respondent wants to impress upon the interviewer that he belongs to a well-to-do family and can afford to spend money on clubs. To obtain facts, the respondents must be conditioned (by good support) to part with the correct facts.

Did u know? **What is the Mode of Collecting the Data?**

The questionnaire can be used to collect information either through personal interview, mail or telephone. The method chosen depends on the information required and also the type of respondent. If the information is to be collected from illiterate individuals, a questionnaire would be the wrong choice.

Type of Questions

Open-ended Questions

These are questions where respondents are free to answer in their own words.

Example: "What factor do you consider while buying a suit"? If multiple choices are given, it could be colour, price, style, brand etc., but some respondents may mention attributes which may not occur to the researcher.

Notes

Therefore, open-ended questions are useful in exploratory research, where all possible alternatives are explored. The greatest disadvantage of open-ended questions is that the researcher has to note down the answer of the respondents verbatim. Therefore, there is a likelihood of the researcher failing to record some information.

Another problem with open-ended question is that the respondents may not use the same frame of reference.

Example: "What is the most important attribute in a job?"

Ans: Pay

The respondent may have meant "basic pay" but interviewer may think that the respondent is talking about "total pay including dearness allowance and incentive". Since both of them refer to pay, it is impossible to separate two different frames.

Dichotomous Question

These questions have only two answers, 'Yes' or 'no', 'true' or 'false' 'use' or 'don't use'.

Do you use toothpaste? Yes …….. No ……

There is no third answer. However sometimes, there can be a third answer:

Example: "Do you like to watch movies?"

Ans: Neither like nor dislike

Dichotomous questions are most convenient and easy to answer.

Close-ended Questions

There are two basic formats in this type:

- Make one or more choices among the alternatives
- Rate the alternatives

Choice among Alternatives

Which of the following words or phrases best describes the kind of person you feel would be most likely to use this product, based on what you have seen in the commercial?

(a) Young …… old …….

Single …… Married ……..

Modern …… Old fashioned ……...

(b) Rating Scale

(I) Please tell us your overall reaction to this commercial?

- A great commercial would like to see again.
- Just so-so, like other commercials.
- Another bad commercial.
- Pretty good commercial.

Notes

(II) Based on what you saw in the commercial, how interested do you feel, you would be buying the products?

- Definitely
- Probably I would buy
- I may or may not buy
- Probably I would not buy
- Definitely I would not buy.

Closed-ended questionnaires are easy to answer. It requires less effort on the part of the interviewer. Tabulation and analysis is easier. There are lesser errors, since the same questions are asked to everyone. The time taken to respond is lesser. We can compare the answer of one respondent to another respondent.

One basic criticism of closed-ended questionnaires is that middle alternatives are not included in this, such as "don't know". This will force the respondents to choose among the given alternative.

Task Which type of questionnaires do you think to be easier to answer? Give reasons to support your argument.

Wordings of Questions

Wordings of particular questions could have a large impact on how the respondent interprets them. Even a small shift in the wording could alter the respondent's answer.

Example:

- "Don't you think that Brazil played poorly in the FIFA cup?" The answer will be 'yes'. Many of them, who do not have any idea about the game, will also most likely say 'yes'. If the question is worded in a slightly different manner, the response will be different.
- "Do you think that, Brazil played poorly in the FIFA cup?" This is a straightforward question. The answer could be 'yes', 'no' or 'don't know' depending on the knowledge the respondents have about the game.
- "Do you think anything should be done to make it easier for people to pay their phone bill, electricity bill and water bill under one roof"?
- "Don't you think something might be done to make it easier for people to pay their phone bill, electricity bill, water bill under one roof"?

A change of just one word as above can generate different responses by respondents.

Guidelines towards the use of correct wording:

Is the vocabulary simple and familiar to the respondents?

Example:

- Instead of using the word 'reasonably', 'usually', 'occasionally', 'generally', 'on the whole'.
- "How often do you go to a movie?" "Often, may be once a week, once a month, once in two months or even more."

Notes

Avoid Double-barreled Questions

These are questions, in which the respondent can agree with one part of the question, but not agree with the other or cannot answer without making a particular assumption.

Example:

- "Do you feel that firms today are employee-oriented and customer-oriented?" There are two separate issues here – [yes] [no]
- "Are you happy with the price and quality of branded shampoo?" [yes] [no]

Avoid Leading and Loading Questions

Leading Questions

A leading question is one that suggests the answer to the respondent. The question itself will influence the answer, when respondents get an idea that the data is being collected by a company. The respondents have a tendency to respond positively.

Example:

- "How do you like the programme on 'Radio Mirchy'? The answer is likely to be 'yes'. The unbiased way of asking is "which is your favorite F.M. Radio station? The answer could be any one of the four stations namely 1. Radio City 2. Mirchy 3. Rainbow 4. Radio-One.
- Do you think that offshore drilling for oil is environmentally unsound? The most probable response is 'yes'. The same question can be modified to eliminate the leading factor.

What is your feeling about the environmental impact of offshore drilling for oil? Give choices as follows:

1. Offshore drilling is environmentally sound.
2. Offshore drilling is environmentally unsound.
3. No opinion.

Loaded Questions

A leading question is also known as a loaded question. In a loaded question, special emphasis is given to a word or a phrase, which acts as a lead to respondent.

Example:

- "Do you own a Kelvinator refrigerator."
- A better question would be "what brand of refrigerator do you own?"
- "Don't you think the civic body is 'incompetent'?"
- Here the word incompetent is 'loaded'.

Are the Questions Confusing?

If there is a question unclear or is confusing, then the respondent becomes more biased rather that getting enlightened.

Notes

Example:"Do you think that the government publications are distributed effectively"?

This is not the correct way, since respondent does not know what the meaning of the word effective distribution is**.** This is confusing. The correct way of asking questions is "Do you think that the government publications are readily available when you want to buy?" Example "Do you think whether value price equation is attractive"? Here, respondents may not know the meaning of **value price** equation.

Applicability

"Is the question applicable to all respondents?" Respondents may try to answer a question even though they don't qualify to do so or may lack from any meaningful opinion.

Example:

1. "What is your present education level"
2. "Where are you working" (assuming he is employed)?
3. "From which bank have you taken a housing loan" (assuming he has taken a loan).

Avoid Implicit Assumptions

An implicit alternative is one that is not expressed in the options. Consider following two questions:

1. Would you like to have a job, if available?
2. Would you prefer to have a job, or do you prefer to do just domestic work?

Even though, we may say that these two questions look similar, they vary widely. The difference is that Q-2 makes explicit the alternative implied in Q-1.

Split Ballot Technique

This is a procedure used wherein (1) The question is split into two halves and (2) Different sequencing of questions is administered to each half. There are occasions when a single version of questions may not derive the correct answer and the choice is not obvious to the respondent.

Example: "Why do you use Ayurvedic soap"? One respondent might say "Ayurvedic soap is better for skin care". Another may say "Because the dermatologist has recommended". A third might say "It is a soap used by my entire family for several years". The first respondent answers the reason for using it at present. The second respondent answers how he started using. The third respondent "the family tradition for using". As can be seen, different reference frames are used. The question may be balanced and rephrased.

Complex Questions

In which of the following do you like to park your liquid funds?

- Debenture
- Preferential share
- Equity linked M.F
- I.P.O
- Fixed deposit

If this question is posed to the general public, they may not know the meaning of liquid fund. Most of the respondents will guess and tick one of them.

Notes

Are the Questions Too Long?

Generally as a thumb rule, it is advisable to keep the number of words in a question not exceeding 20. The question given below is too long for the respondent to comprehend, leave alone answer.

Example: Do you accept that the people whom you know, and associate yourself have been receiving ESI and P.F benefits from the government accept a reduction in those benefits, with a view to cut down government expenditure, to provide more resources for infrastructural development?

Yes _____ No_____ Can't say_____

Participation at the Expense of Accuracy

Sometimes the respondent may not have the information that is needed by the researcher.

Example:

- The husband is asked a question "How much does your family spend on groceries in a week"? Unless the respondent does the grocery shopping himself, he will not know how much has been spent. In a situation like this, it will be helpful to ask a 'filtered question'. An example of a filtered question can be, "Who buys the groceries in your family"?
- "Do you have the information of Mr. Ben's visit to Bangalore"? Not only should the individual have the information but also s(he) should remember the same. The inability to remember the information is known as "recall loss".

Sequence and Layout

Some guidelines for sequencing the questionnaire are as follows:

Divide the questionnaire into three parts:

(1) Basic information (2) Classification (3) Identification information. Items such as age, sex, income, education etc. are questioned in the classification section. The identification part involves body of the questionnaire. Always move from general to specific questions on the topic. This is known as funnel sequence. Sequencing of questions is illustrated below:

(1) Which TV shows do you watch?

Sports ..... News .....

(2) Which among the following are you most interested in?

Sports ..... News .....

Music ..... Cartoon .....

(3) Which show did you watch last week?

World Cup Football .....

Bournvita Quiz Contest .....

War News in the Middle East .....

Tom and Jerry cartoon show .....

Notes

The above three questions follow a funnel sequence. If we reverse the order of question and ask "which show was watched last week"?, the answer may be biased. This example shows the importance of sequencing.

Layout: How the questionnaire looks or appears.

Example: Clear instructions, gaps between questions, answers and spaces are part of layout. Two different layouts are shown below:

Layout – 1 **How old is your bike?**

_____Less than 1 year _____1 to 2 years _____2 to 4 years _____ more than 4 years.

Layout – 2 **How old is your bike?**

_____Less than 1 year

_____1 to 2 years.

_____2 to 4 years.

_____ More than 4 years.

From the above example, it is clear that layout – 2 is better. This is because likely respondent error due to confusion is minimised.

Therefore, while preparing a questionnaire start with a general question. This is followed by a direct and simple question. This is followed by more focused questions. This will elicit maximum information.

Forced and Unforced Scales

Suppose the questionnaire is not provided with 'don't know' or 'no option', then the respondent is forced to choose one side or the other. A 'don't know' is not a neutral response. This may be due to genuine lack of knowledge.

Balanced and Unbalanced Scales

In a balanced scale, the numbers of favorable responses are equal to the number of unfavorable responses. If the researcher knows that there is a possibility of a favourable response, it is best to use unbalanced scale.

Use Funnel Approach

Funnel sequencing gets the name from its shape, starting with broad questions and progressively narrowing down the scope. Move from general to specific examples.

1. How do you think this country is getting along in its relations with other countries?
2. How do you think we are doing in our relations with the US?
3. Do you think we ought to be dealing with US?
4. If yes, what should be done differently?
5. Some say we are very weak on the nuclear deal with the US, while, some say we are OK. What do you feel?

The first question introduces the general subject. In the next question, a specific country is mentioned. The third and fourth questions are asked to seek views. The fifth question is to seek a specific opinion.

Notes

Pre-testing of Questionnaire

Pre-testing of a questionnaire is done to detect any flaws that might be present.

Example: The word used by researcher must convey the same meaning to the respondents. Are instructions clear skip questions clear?

One of the prime conditions for pre-testing is that the sample chosen for pre-testing should be similar to the respondents who are ultimately going to participate. Just because a few chosen respondents fill in all the questions going does not mean that the questionnaire is sound.

How many Questions to be Asked?

The questionnaire should not be too long as the response will be poor. There is no rule to decide this. However, the researcher should consider that if he were the respondent, how he would react to a lengthy questionnaire. One way of deciding the length of the questionnaire is to calculate the time taken to complete the questionnaire. He can give the questionnaire to a few known people to seek their opinion.

Task Give one example for each of the following type of the questions:

1. Leading question
2. Double-barreled question
3. Close-ended question
4. Fixed alternative question
5. Split-ballot question

7.3.5 Mail Questionnaire

Mail questionnaires can be explained as the questionnaires that are mailed to the respondents who can complete them at their convenience in their homes and at their own pace. They are expected to meet with a better response rate when respondents are notified in advance about the forthcoming survey and a reputed research organization administers them with its own introductory cover letter.

Advantages of Mail Questionnaire

1. Easier to reach a larger number of respondents throughout the country.
2. Since the interviewer is not present face to face, the influence of interviewer on the respondent is eliminated.
3. Where the questions asked are such that they cannot be answered immediately, and needs some thinking on the part of the respondent, the respondent can think over leisurely and give the answer.
4. Saves cost (cheaper than interview).
5. No need to train interviewers.
6. Personal and sensitive questions are well answered.

Notes

Limitations of Mail Questionnaire

1. It is not suitable when questions are difficult and complicated.

Example: "Do you believe in value price relationship"?

2. When the researcher is interested in a spontaneous response, this method is unsuitable. Because thinking time allowed to the respondent will influence the answer.

Example: "Tell me spontaneously, what comes to your mind if I ask you about cigarette smoking".

3. In case of a mail questionnaire, it is not possible to verify whether the respondent himself/ herself has filled the questionnaire. If the questionnaire is directed towards the housewife, say, to know her expenditure on kitchen items, she alone is supposed to answer it. Instead, if her husband answers the questionnaire, the answer may not be correct.

4. Any clarification required by the respondent regarding questions is not possible.

Example: Prorated discount, product profile, marginal rate, etc., may not be understood by the respondents.

5. If the answers are not correct, the researcher cannot probe further.

6. Poor response (30%) - Not all reply.

Additional Consideration for the Preparation of Mail Questionnaire

1. It should be shorter than the questionnaire used for a personal interview.

2. The wording should be extremely simple.

3. If a lengthy questionnaire has to be made, first write a letter requesting the cooperation of the respondents.

4. Provide clear guidance, wherever necessary.

5. Send a pre-addressed and stamped envelope to receive the reply.

Task A nationalised bank wants to determine what is most effective way to increase responses to their mail questionnaire? Three possibilities are:

1. Issue a gift coupon for ₹ 25 so that the respondent can go to a specified store to avail the gift item.

2. Ask the respondent to note their name and address in the completed questionnaire. Thereafter they will be mailed a cheque for ₹ 50.

3. Along with a questionnaire, write a letter stating pen set as gifts would be sent to them, after receiving the completed questionnaire. Mail the questionnaire to 2,000 respondents chosen from four metros. Set up an experiment in which the above incentives can be tested and the most appropriate method identified.

Notes

7.3.6 Schedule Method

Schedule may be defined as a proforma that contains a set of questions which are asked and filled by an interviewer in a face to face situation with another. It is a standardized device or tool of observation to collect the data in an objective manner. In this method of data collection the interviewer puts certain questions and the respondent furnishes certain answers and the interviewer records as they are given.

Purpose/Objectives of the Schedule

The main objectives of the schedule are as follows:

- ***Delimitation of the topic:*** A schedule is always about a definite item of enquiry. It's subject is a single and isolated item rather than the research subject in general. The schedule therefore delimits and specifies the subject of enquiry.
- ***Aids to Memorize:*** It is not possible for the interviewer to keep in mind or memorize all the information that he collects from different respondents. Without a standardized tool, he might ask different questions to different respondents and thereby get confused when he requires to analyze and tabulate the data. Therefore schedule acts as an "aide memoire".
- ***Aid to classification and analysis:*** Another objective of the schedule is to tabulate and analyze the data collected in a scientific and homogeneous manner.

Types of Schedules

These are as follows:

- ***Observation Schedule:*** The schedules which are used for observation are known as observation schedules. Using this schedule, observer records the activities and responses of an individual respondent or a group of respondents under specific conditions. The main purpose of the observation schedule is to verify information.
- ***Rating Schedule:*** Rating schedules are used to assess the attitudes, opinions, preferences, inhibitions, perceptions and other similar elements or attributes of respondent. Such measurement is done using a Rating Scale.
- ***Document Schedule:*** These schedules are used in exploratory research to obtain data regarding written evidence and case histories from autobiography, diary, or records of government etc. It is an important method for collecting preliminary data or for preparing a source list.
- ***Institution Survey Schedules:*** This type of schedule is used for studying different problems of institutions.
- ***Interview Schedule:*** Using his schedule, an interviewer presents the questions to the interviewee and records his responses in the given space of the questionnaire.

Merits of Schedule Method

The schedule method has the following merits:

- ***Higher response:*** In the schedule, since a research worker is present and he can explain and persuade the respondent, response rate is high. In case of any mistake in the schedule, the researcher can rectify it.

Notes

- ***Saving of time:*** While filling the schedule, the researcher may use abbreviation or short forms for answers, he may also generate a template. All these steps help in saving of time in data collection.
- ***Personal contact:*** In the schedule method there is a personal contact between the respondent and the field worker. The behaviour, and character of respondent obviously facilitates the research work.
- ***Human touch:*** Sometimes reading something does not impress as much as when the same is heard or spoken by experts as they are able to lay the right emphasis. This greatly improves the response.
- ***Deeper probe:*** Through this method it is possible to probe deeper into the personality, living conditions, values, etc., of the respondents.
- ***Defects in sampling are detected:*** If there are some defects in schedule during sampling it easily come to the notice and can be rectified by the researcher.
- ***Removal of doubts:*** Presence of enumerator removes the doubts in the minds of respondent on the one hand and avoid from the respondent artificial replies owing to fear of cross checking on the other hand.
- ***Human elements make the study more reliable and dependable:*** The presence of human elements makes the situation more attractive and interesting which helps in making interview useful and reliable.

Limitations of the Schedule Method

Following are the main limitation of the schedule method:

- ***Costly and time-consuming:*** This method is costly and time consuming due to its basic requirement of interviewing the respondents. This becomes a serious limitation when respondents are not found in a particular region but are scattered over a wide area.
- ***Need of trained field workers:*** The schedule method requires involvement of well trained and experienced field workers. This involves great cost and sometimes workers are not easily available forcing engagement of inexperienced hands, which defeats the purpose trained of research.
- ***Adverse effect of personal presence:*** Sometimes personal presence of enumerator becomes an inhibiting factor. Many people despite knowing certain facts cannot say them in the presence of others.
- ***Organizational difficulties:*** If the field of research is dispersed, it becomes difficult to organize it. Getting trained manpower, assigning them duties and then administrating the research is a very difficult task.

Characteristics of a Good Schedule

The following are the essentials or characteristics of a good schedule:

- ***Accurate communication:*** It means that the questions given in the schedule should enable the respondent to understand the context in which they are asked.
- ***Accurate response:*** The schedule should structure in such a manner so that the required information are accurate and secured. For this, following steps should be taken.
    - The size of the schedule should be precise and attractive.

Notes

- The questions should be clearly worded and should be unambiguous.
- The questions should be free from any subjective evaluation.
- Questions should be inter-linked.
- Information sought should be capable of tabulation and subsequent statistical analysis.

Suitability of Schedule Method

This method is generally applied in the following situations:

- The field of investigation is wide and dispersed.
- Where the researcher requires quick result at low cost.
- Where the respondents are educated.
- Where trained and educated investigators are available.

7.3.7 Sample Questionnaires

A Study of Customer Retention as Adopted by Textile Retail Outlets

Note: Information gathered will be strictly confidential. We highly appreciate your cooperation in this regard.

1. Name of the outlet:
2. Address:
3. Do you have regular customers?

   Yes [ ] No [ ]
4. How often do your regular customers visit your outlet?

   Weekly [ ] Once in a month [ ] Twice in a month [ ]

   Once in 2 months [ ] 2-3 months [ ] Once in 6 months [ ]
5. Do you maintain any records of your regular customers?

   Yes [ ] No [ ]
6. What percentage of your customers are regular? % [ ]
7. Do you think that we can use the above as a retention strategy of customers for your outlets?

   Yes [ ] No [ ]
8. What are the different products that you handle in your outlets?

   Formals [ ] Casuals/Kids wear [ ] Ladies dress materials [ ]

   Sarees [ ] Others (Specify) – – – – – – – – – –
9. What type of customers (socio-economic) visit your outlets?

   Low income [ ] Middle income [ ] High income [ ]
10. Why do you think they come to your outlet?

    Product variety [ ] Price discount [ ] Easy gain to products [ ]

Contd...

Notes

Parking facility [ ] Store layout [ ] Quality [ ]

Reasonable price [ ] Others (Specify)

11. Rank the factors that influence customers to visit your outlet:

Credit facility [ ] Price discount [ ] Gifts [ ]

Easy gain to products [ ] Parking facility [ ]

Store layout [ ] Product variety [ ]

Quality and reasonable price [ ] Others (Specify) ————————————

12. What do customers expect from the retail outlet?

Credit facility [ ] Gift coupon [ ] Price discount [ ]

Price reduction easy accessibility of product [ ]

Quality and reasonable price [ ]

Others (Specify) ————————————-

13. Do you have any retention strategy adopted to keep in touch with the customer?

Gifts on special occasion

(a) Birthday gift [ ]

(b) Anniversary gift [ ]

(c) Festivals/Customer relationship [ ]

(d) Others (Specify) ————————

14. Which one do you think is most effective? Please rank them?

(a) Birthday gift [ ]

(b) Anniversary gift [ ]

(c) Festivals/Customer relationship [ ]

(d) Others (Specify) ————————————

Thanking you for Sparing your Valuable Time

A Study on Customer Preferences of P.C.

Date: Place:

Form No: [ ] [ ] [ ] [ ] [ ]

1. Personal Profile

(a) Name: [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ]

(b) Address: [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ]

[ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ]

[ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ]

(c) Sex: Male [ ] Female [ ]

(d) Age: [ ] [ ] years

(e) Occupation: Self-employed [ ] Professional [ ] Service [ ] Housewife [ ]

Contd...

Notes

2. Do you own a P.C? Yes [ ] No [ ]

   (a) If yes, whether: branded [ ] unbranded [ ]

   (b) If no, do you plan to buy one? Near future [ ] Distant future [ ] Can't say [ ]

   (Less than 6 months) (Less than a year)

   If so, whether: branded [ ] unbranded [ ]

3. What is the utility of the PC to you?

Education	[ ]	Business	[ ]
Infotainment	[ ]	Internet/Communication	[ ]

4. What is the most important factor that matters while buying a PC?

Quality	[ ]	Price	[ ]
Service	[ ]	Finance facility	[ ]

5. Before deciding on the vendor, which factor goes into your consideration?

Vendor's Reputation	[ ]	Technical Expertise	[ ]
Client Base	[ ]		

6. How did you come to know about the vendor?

Friendly / Family	[ ]	Press Ads	[ ]
Direct Mailers	[ ]	Reference Website	[ ]

7. Which configuration would you decide on while buying a PC?

Standard	[ ]	Intermediate	[ ]
Latest / Advanced	[ ]		

8. In your PC, would you prefer?

Conventional Design	[ ]	Innovative Design	[ ]

If new, why:

New design distracts attention -

New design means increased price -

New design is hard to adapt -

If Innovative, why?

To create own identity

Out of business need -

Space management -

9. Rate the following four factors important for innovative design, starting with the most preferred:

(a)	Size	(b)	Shape
(c)	Colour/ordinary	(d)	Portability and Sturdiness
1.	————————————	3.	————————————
2.	————————————	4.	————————————

Contd...

Notes

10. To what extent would the computer increase your efficiency?

Negligible	[ ]	20-40%	[ ]
40-60%	[ ]	More	[ ]

11. How many hours on an average per week would you use your PC?

0 to 5 hours	[ ]	6 to 12 hours	[ ]
13 to 18 hours	[ ]	More	[ ]

12. While using your PC, most of the time would be for:

Education	[ ]	Accounting	[ ]
Net surfing	[ ]	Correspondence	[ ]

13. Remarks _____

_____

_____

_____

Signature of the Respondent _____

Self Assessment

Fill in the blanks:

9. A ……. is a research instrument consisting of a series of questions and other prompts for the purpose of gathering information from respondents.
10. The main objective of …….. is to conceal the topic of enquiry by using a disguised stimulus.
11. ……. are questions where respondents are free to answer in their own words.
12. A ……. question is one that suggests the answer to the respondent.

7.4 Summary

- Data sources are broadly classified into primary and secondary data.
- Primary data is one which is collected by the investigator himself for the purpose of a specific inquiry or study. The data directly collected by the researcher, with respect to the problem under study, is known as primary data.
- Observation method has a limitation i.e., certain attitudes, knowledge, motivation etc. cannot be measured by this method.
- Secondary data are statistics that already exists. These may not be readily used because these data are collected for some other purpose.
- There are two types of secondary data (1) Internal and (2) External secondary data.
- Census is the most important among secondary data.
- Syndicated data is an important form of secondary data which may be classified into (a) Consumer purchase data (b) Retailer and wholesaler data (c) Advertising data.
- Questionnaire can be administered either in person or on-line or Mail questionnaire.

Notes

- Questions in a questionnaire may be classified into open question, close ended questions, dichotomous questions etc.

7.5 Keywords

Closed-ended questions: There are two basic formats in this type: (a) Make one or more choices among the alternatives and (b) Rate the alternatives.

Dichotomous questions: These questions have only two answers, 'Yes' or 'no', 'true' or 'false' 'use' or 'don't use'.

Internal Secondary Data: Is that data which is a part of company's record, for which research is already conducted.

Leading question: A leading question is one that suggests the answer to the respondent.

Open-ended questions: These are questions where respondents are free to answer in their own words.

Primary Data: Data directly collected by the researcher, with respect to problem under study, is known as primary data.

Recency: This refers to "How old is the information?" If it is five years old, it may be useless.

Structured disguised Questionnaire: This type of Questionnaire is used to find, peoples' attitude, when a direct undisguised question produces a bias.

7.6 Review Questions

1. What is primary data?
2. What are the various methods available for collecting primary data?
3. What are the several methods used to collect data by observation method?
4. What are the advantages and limitations of collecting data by observation method?
5. What would you define as the characteristics of a good questionnaire?
6. By the help of examples only, explain what is meant by leading/loading question?

Answers: Self Assessment

1. Primary data
2. Authenticity
3. Internal secondary data
4. Interview
5. Observation, questioning
6. Direct observation
7. Qualitative research
8. Depth
9. Questionnaire
10. Non-Structured and Disguised Questionnaire

Notes

11. Open-ended Questions
12. Leading

7.7 Further Readings

Books

Boyd, Westfall, and Stasch, *Marketing Research - Text and Cases*, All India Traveller Bookseller, New Delhi.

Brown, F.E., *Marketing Research - A Structure for Decision-making*, Addison-Wesley Publishing Company.

Kothari, C.R., *Research Methodology - Methods and Techniques*, Wiley Eastern Ltd.

S.N. Murthy and U. Bhojanna, *Business Research Methods*, Excel Books, 2007.

Stockton and Clark, *Introduction to Business and Economic Statistics*, D.B. Taraporevala Sons and Co. Private Limited, Bombay.

Online links

www.indiastudychannel.com

www.scribd.com/doc

www.soas.ac.uk

www.web-source.net

Notes

Unit 8: Sampling and Sampling Distribution

CONTENTS

Objectives
Introduction
8.1 Meaning of Sampling
8.1.1 Sample Frame
8.1.2 When is a Census Appropriate?
8.1.3 When is Sample Appropriate?
8.2 Sampling Process
8.3 Types of Sample Design
8.3.1 Probability Sampling Techniques
8.3.2 Non-probability Sampling Techniques
8.4 Distinction between Probability Sample and Non-probability Sample
8.5 Errors in Sampling
8.5.1 Sampling Error
8.5.2 Non-sampling Error
8.5.3 Sampling Frame Error
8.5.4 Non-response Error
8.5.5 Data Error
8.5.6 Failure of the Interviewer to Follow Instructions
8.6 Sample Size Decision
8.7 Sampling Distribution
8.8 Summary
8.9 Keywords
8.10 Review Questions
8.11 Further Readings

Objectives

After studying this unit, you will be able to:

- Define sampling;
- Identify the steps involved in the sampling process;
- Discuss the types of sampling design;
- Summaries the probability and non-probability sampling;
- Explain the types of errors in sampling;
- Interpret the sampling size;
- Explain the sampling distribution.

Notes

Introduction

The most important task in carrying out a survey is to select the sample. Sample selection is undertaken for practical impossibility to survey the population. By applying rationality in selection of samples, we generalise the findings of our research.

In carrying out a survey relating to the research, we should first select the problem and study its implications in different areas. Selection of the research problem, as has already been stated, should be in line with the researcher's interest, chain of thinking and existing research in the same area and should have some direct utility. What is most important in selecting a research problem is that the research topic should be within manageable limits.

8.1 Meaning of Sampling

Sampling is the process of selecting units (e.g., people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen. Each observation measures one or more properties (weight, location, etc.) of an observable entity enumerated to distinguish objects or individuals. Survey weights often need to be applied to the data to adjust for the sample design. Results from probability theory and statistical theory are employed to guide practice. A sample is a part of a target population, which is carefully selected to represent the population. Sampling frame is the list of elements from which the sample is actually drawn. Actually, sampling frame is nothing but the correct list of population.

8.1.1 Sample Frame

Sampling frame is the list of elements from which the sample is actually drawn. Actually, sampling frame is nothing but the correct list of population.

Example: Telephone directory, Product finder, Yellow pages.

Did u know? **What is the distinction between census and sampling?**

Census refers to complete inclusion of all elements in the population. A sample is a sub-group of the population.

8.1.2 When is a Census Appropriate?

1. A census is appropriate if the size of population is small.

Example: A researcher may be interested in contacting firms in iron and steel or petroleum products industry. These industries are limited in number, so a census will be suitable.

2. Sometimes, the researcher is interested in gathering information from every individual.

Example: Quality of food served in a mess.

8.1.3 When is Sample Appropriate?

1. When the size of population is large.
2. When time and cost are the main considerations in research.

Notes

3. If the population is homogeneous.
4. Also, there are circumstances when a census is not possible.

Example: Reactions to global advertising by a company.

8.2 Sampling Process

Sampling process consists of seven steps. They are:

1. Define the population.
2. Identify the sampling frame.
3. Specify the sampling unit.
4. Selection of sampling method.
5. Determination of sample size.
6. Specify sampling plan.
7. Selection of sample.

(1) ***Define the population:*** Population is defined in terms of:

(1) Elements

(2) Sampling units

(3) Extent

(4) Time.

Example: If we are monitoring the sale of a new product recently introduced by a company, say (shampoo sachet) the population will be:

1. Element - Company's product
2. Sampling unit - Retail outlet, super market
3. Extent - Hyderabad and Secunderabad
4. Time - April 10 to May 10, 2006

(2) ***Identify the sampling frame:*** Sampling frame could be (a) Telephone Directory (b) Localities of a city using the municipal corporation listing (c) Any other list consisting of all sampling units.

Example: You want to learn about scooter owners in a city. The RTO will be the frame, which provides you names, addresses and the types of vehicles possessed.

(3) ***Specify the sampling unit:*** Individuals who are to be contacted are the sampling units. If retailers are to be contacted in a locality, they are the sampling units.

Sampling unit may be husband or wife in a family. The selection of sampling unit is very important. If interviews are to be held during office timings, when the heads of families and other employed persons are away, interviewing would under-represent employed persons, and over-represent elderly persons, housewives and the unemployed.

(4) ***Selection of sampling method:*** This refers to whether (a) probability or (b) non-probability methods are used.

Notes

(5) ***Determine the sample size:*** This means we need to decide "how many elements of the target population are to be chosen?" The sample size depends upon the type of study that is being conducted.

Example: If it is an exploratory research, the sample size will be generally small. For conclusive research, such as descriptive research, the sample size will be large.

The sample size also depends upon the resources available with the company. It depends on the accuracy required in the study and the permissible errors allowed.

(6) ***Specify the sampling plan:*** A sampling plan should clearly specify the target population. Improper defining would lead to wrong data collection.

Example: This means that, if a survey of a household is to be conducted, a sampling plan should define a "household" i.e., "Does the household consist of husband or wife or both", minors etc., "Who should be included or excluded". Instructions to the interviewer should include "How he should obtain a systematic sample of households, probability sampling non–probability sampling". Advise him on what he should do to the household, if no one is available.

(7) ***Select the sample:*** This is the final step in the sampling process.

Self Assessment

Fill in the blanks:

1. ..... is the process of selecting units from a population of interest.
2. A sample is a part of a ..... population.
3. Sampling ..... is the list of elements from which the sample is actually drawn.
4. A census is appropriate if the size of population is ..... .
5. A sampling plan should clearly specify the ..... population.
6. The sample size depends upon the ..... available with the company.

8.3 Types of Sample Design

Sampling is divided into two types:

- ***Probability sampling:*** In a probability sample, every unit in the population has equal chances for being selected as a sample unit.
- ***Non–probability sampling:*** In the non-probability sampling, the units in the population have unequal or negligible, almost no chances for being selected as a sample unit.

8.3.1 Probability Sampling Techniques

1. Random sampling.
2. Stratified random sampling.
3. Systematic sampling.
4. Cluster sampling.
5. Multi-stage sampling.

Notes

Random Sampling

Simple random sample is a process in which every item of the population has an equal probability of being chosen.

There are two methods used in the random sampling:

(1) Lottery method

(2) Using random number table.

(1) ***Lottery method:*** Take a population containing four departmental stores: A, B, C and D. Suppose we need to pick a sample of two stores from the population using a simple random procedure. We write down all possible samples of two. Six different combinations, each containing two stores from the population, are AB, AD, AC, BC, BD, CD. We can now write down six sample combination on six identical pieces of paper, fold the piece of paper so that they cannot be distinguished. Put them in a box. Mix them and pull one at random. This procedure is the lottery method of making a random selection.

(2) ***Using random number table:*** A random number table consists of a group of digits that are arranged in random order, i.e., any row, column, or diagonal in such a table contains digits that are not in any systematic order. There are three tables for random numbers (a) Tippet's table (b) Fisher and Yate's table (c) Kendall and Raington table.

The table for random number is as follows:

40743	39672
80833	18496
10743	39431
88103	23016
53946	43761
31230	41212
24323	18054

Example: Taking the earlier example of stores. We first number the stores.

1 A 2 B 3 C 4 D

The stores A, B, C and D have been numbered as 1, 2, 3 and 4.

We proceed as follows, in order to select two shops out of four randomly:

Suppose, we start with the second row in the first column of the table and decide to read diagonally. The starting digit is 8. There are no departmental stores with the number 8 in the population. There are only four stores. Move to the next digit on the diagonal, which is 0. Ignore it, since it does not correspond to any of the stores in the population. The next digit on the diagonal is 1 which corresponds to store A. Pick A and proceed until we get two samples. In this case, the two departmental stores are 1 and 4. The sample derived from this consists of departmental stores A and D.

In random sampling, there are two possibilities (1) Equal probability (2) Varying probability.

Equal Probability

This is also called as the random sampling with replacement.

Notes

Example: Put 100 chits in a box numbered 1 to 100. Pick one number at random. Now the population has 99 chits. Now, when a second number is being picked, there are 99 chits. In order to provide equal probability, the sample selected is being replaced in the population.

Varying Probability

This is also called random sampling without replacement. Once a number is picked, it is not included again. Therefore, the probability of selecting a unit varies from the other.

In our example, it is 1/100, 1/99, 1/98, 1/97 if we select four samples out of 100.

Systematic Random Sampling

There are three steps:

(1) Sampling interval K is determined by the following formula:

$$K = \frac{\text{No. of units in the population}}{\text{No. of units desired in the sample}}$$

(2) One unit between the first and Kth unit in the population list is randomly chosen.

(3) Add Kth unit to the randomly chosen number.

Example: Consider 1,000 households from which we want to select 50 units.

$$\text{Calculate } K = \frac{1000}{50} = 20$$

To select the first unit, we randomly pick one number between 1 to 20, say 17. So our sample begins with 17,37,57…….. Please note that only the first item was randomly selected. The rest are systematically selected. This is a very popular method because we need only one random number.

Stratified Random Sampling

A probability sampling procedure in which simple random sub-samples are drawn from within different strata, which are, more or less equal on some characteristics. Stratified sampling are of two types:

1. ***Proportionate stratified sampling:*** The number of sampling units drawn from each stratum is in proportion to the population size of that stratum.
2. ***Disproportionate stratified sampling:*** The number of sampling units drawn from each stratum is based on the analytical consideration, but not in proportion to the size of the population of that stratum.

Sampling process is as follows:

1. The population to be sampled is divided into groups (stratified).
2. A simple random sample is chosen.

Notes

Notes **Reason for Stratified Sampling**

Sometimes, marketing professionals want information about the component part of the population. Assume there are three stores. Each store forms a strata and the sampling from within each strata is being selected. The resultant might be used to plan different promotional activities for each store strata.

Suppose a researcher wishes to study the retail sales of products, such as tea in a universe of 1,000 grocery stores (Kirana shops included). The researcher can first divide this universe into three strata based on the size of the store. This benchmark for size could be only one of the following (a) floor space (b) volume of sales (c) variety displayed etc.

Size of stores	No. of stores	Percentage of stores
Large stores	2,000	20
Medium stores	3,000	30
Small stores	5,000	50
	10,000	**100**

Suppose we need 12 stores, then choose four from each strata, at random. If there was no stratification, simple random sampling from the population would be expected to choose two large stores (20% of 12) about four medium stores (30% of 12) and about six small stores (50% of 12).

As can be seen, each store can be studied separately using the stratified sample.

Stratified sampling can be carried out with:

1. Same proportion across the strata proportionate stratified sample.
2. Varying proportion across the strata disproportionate stratified sample.

Example:

Size of stores	No. of stores(Population)	Sample Proportionate	Sample Disproportionate
Large	2,000	20	25
Medium	3,000	30	35
Small	5,000	50	40

Estimation of universe mean with a stratified sample.

Example:

Size of stores	Sample Mean Sales per store	No. of stores	Percent of stores
Large	200	2000	20
Medium	80	3000	30
Small	40	5000	50
		10,000	**100**

The population mean of monthly sales is calculated by multiplying the sample mean by its relative weight.

$200 \times 0.2 + 80 \times 0.3 + 40 \times 0.5 = 84$

Notes

Sample Proportionate

If N is the size of the population.

n is the size of the sample.

i represents 1, 2, 3,……..k [number of strata in the population]

∴ Proportionate sampling

$$p = \frac{n_1}{N_1} = \frac{n_2}{N_2} = ... = \frac{n_k}{N_k} = \frac{n}{N}$$

$$\frac{n_1}{N_1} = \frac{n}{N} = n_1 = \frac{n}{N} \times n_1 \text{ and so on}$$

n_1 is the sample size to be drawn from stratum 1

$n_1 + n_2 + \ldots\ldots\ldots\ldots n_k = n$ [Total sample size of the all strata]

Example: A survey is planned to analyse the perception of people towards their own religious practices. The population consists of various religions, viz., Hindu, Muslim, Christian, Sikh, Jain, assuming a total of 10,000. Hindu, Muslim, Christian, Sikh and Jains consists of 6,000, 2,000, 1,000, 500 and 500 respectively. Determine the sample size of each stratum by applying proportionate stratified sampling, if the sample size required is 200.

Solution: Total population, N=10,000

Population in the strata of Hindus N_1 = 6,000

Population in the strata of Muslims N_2 = 2,000

Population in the strata of Christians N_3 = 1,000

Population in the strata of Sikhs N_4 = 500

Population in the strata of Jains N_5 = 500

Proportionate Stratified Sampling

$$P = \frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} = \frac{n_4}{N_4} = \frac{n_5}{N_5} = \frac{n}{N}$$

∴ Let us determine the sample size of strata N_1

$$\frac{n_1}{N_1} = \frac{n}{N} \times N_1 = \frac{200}{10,000} \times 6,000$$

$$= 20 \times 6$$

$$= 120$$

$$n_2 = \frac{n}{N} \times N_2 = \frac{200}{10,000} \times 2,000$$

$$= 40$$

Notes

$$n_3 = \frac{n}{N} \times N_3 = \frac{200}{10,000} \times 1,000$$

$$= 20$$

$$n_4 = \frac{n}{N} \times N_4 = \frac{200}{10,000} \times 500$$

$$= 10$$

$$n_5 = \frac{n}{N} \times N_5 = 10$$

$$n = n_1 + n_2 + n_3 + n_4 + n_5$$

$$= 120 + 40 + 20 + 10 + 10$$

$$= 200$$

Sample Disproportion

Let is the variance of the stratum i,

where $i = 1,2,3\ldots\ldots\ldots k.$

The formula to compute the sample size of the stratum i is the variance of the stratum i,

where size of stratum i

r_i = Sample size of stratum i

$$r_i = \frac{Ni}{N}$$

r_i = Ratio of the size of the stratum I with that of the population.

N_i = Population of stratum i

N = Total population.

Example: The Government of India wants to study the performance of women self help groups (WSHGs) in three regions viz. North, South and West. The total number of WSHGs is 1,500. The number of groups in North, South and West are 600, 500 and 400 respectively. The Government found more variation between WSHGs in the North, South and West regions. The variance of performance of WSHGs in these regions are 64, 25 and 16 respectively. If the disappropriate stratified sampling is to be used with the sample size of 100, determine the number of sampling units for each region.

Solutions:

Total Population $N = 1,500$

Size of the stratum 1, $N_1 = 600$

Size of the stratum 2, $N_2 = 500$

Size of the stratum 3, $N_3 = 400$

Variance of stratum 1, $\sigma = 12 = 64$

Notes

Variance of stratum 2, $\sigma = 22 = 25$

Variance of stratum 3, $\sigma = 32 = 16$

Sample size n = 100

Stratum Number	Size of the stratum N_i	$r_i = \frac{N_i}{N}$	σ_i	$r_i\lambda_{in}$	$r_i\sigma_{in} = \frac{r_i\sigma_{in}}{\sum_1^3 r_i\sigma_i}$
1	600	0.4	8	3.2	54
2	500	0.33	5	1.65	28
3	400	0.26	4	1.04	18
Total					100

Cluster Sampling

The following steps are followed:

1. The population is divided into clusters.
2. A simple random sample of few clusters is selected.
3. All the units in the selected cluster are studied.

Step 1: The above mentioned cluster sampling is similar to the first step of stratified random sampling. But the two sampling methods are different. The key to cluster sampling is decided by how homogeneous or heterogeneous the clusters are.

A major advantage of simple cluster sampling is the case of sample selection. Suppose, we have a population of 20,000 units from which we wish to select 500 units. Choosing a sample of that size is a very time-consuming process, if we use Random Numbers table. Suppose, the entire population is divided into 80 clusters of 250 units each, we can choose two sample clusters (2 × 250=500) easily by using cluster sampling. The most difficult job is to form clusters. In marketing, the researcher forms clusters so that he can deal with each cluster differently.

Example: Assume there are 20 households in a locality.

Cross			Houses	
1	X_1	X_2	X_3	X_4
2	X_5	X_6	X_7	X_8
3	X_9	X_{10}	X_{11}	X_{12}
4	X_{13}	X_{14}	X_{15}	X_{16}

We need to select eight houses. We can choose eight houses at random. Alternatively, two clusters, each containing four houses can be chosen. In this method, every possible sample of eight houses would have a known probability of being chosen – i.e. chance of one in two. We must remember that in the cluster, each house has the same characteristics. With cluster sampling, it is impossible for certain random sample to be selected. For example, in the cluster sampling process described above, the following combination of houses could not occur: X_1, X_2, X_5, X_6, X_9, X_{10}, X_{13} and X_{14}. This is because the original universe of 16 houses have been redefined as a universe of four clusters. So only clusters can be chosen as a sample.

Notes

Example: Suppose, we want to have 7500 households from all over the country. In such a case, from the first stage, District, say 30 districts out of 600 are selected from all over the country.

I Stage - Cities: Suppose 5 cities are selected out of each 30 districts; and

II Stage - Wards/Localities: say 10 wards/localities are selected from each city

III Stage - Households: 50 households are selected from each ward/locality.

In stage I, we can employ stratified sampling

In stage II, we can use cluster sampling

In stage III, we can have simple random sampling.

Caution The use of various methods shall give individually contribute towards accuracy, cost, time, etc. This leads us to conclude that multistage sampling leads to saving of time, labour and money. Apart from this wherever an appropriate frame is not available, the use of multistage sampling has universal appeal.

Multi-stage Sampling

The name implies that sampling is done in several stages. This is used with stratified/cluster designs.

An illustration of double sampling is as follows.

The management of a newly-opened club is solicits new membership. During the first rounds, all corporate were sent details so that those who are interested may enroll. Having enrolled, the second round concentrates on how many are interested to enroll for various entertainment activities that club offers such as billiards, indoor sports, swimming, and gym etc. After obtaining this information, you might stratify the interested respondents. This will also tell you the reaction of new members to various activities. This technique is considered to be scientific, since there is no possibility of ignoring the characteristics of the universe.

Task What are the advantages and disadvantages of multi-stage sampling? Enlist.

Area Sampling

This is a Probability sampling, a special form of cluster sampling.

Example: If someone wants to measure the sales of toffee in retail stores, one might choose a city locality and then audit toffee sales in retail outlets in those localities.

The main problem in area sampling is the non-availability of lists of shops selling toffee in a particular area. Therefore, it would be impossible to choose a probability sample from these outlets directly. Thus, the first job is to choose a geographical area and then list out outlets selling toffee. Then follows the probability sample for shops among the list prepared.

Example: You may like to choose shops which sell the brand—Cadbury dairy milk. The disadvantage of the area sampling is that it is expensive and time-consuming.

Notes

8.3.2 Non-probability Sampling Techniques

1. Deliberate sampling
2. Shopping Mall Intercept Sampling
3. Sequential sampling
4. Quota sampling
5. Snowball sampling
6. Panel samples

Deliberate or Purposive Sampling

This is also known as the judgment sampling. The investigator uses his discretion in selecting sample observations from the universe. As a result, there is an element of bias in the selection. From the point of view of the investigator, the sample thus chosen may be a true representative of the universe. However, the units in the universe do not enjoy an equal chance of getting included in the sample. Therefore, it cannot be considered a probability sampling.

Example: Test market cities are being selected, based on the judgment sampling, because these cities are viewed as typical cities matching with certain demographical characteristics.

Shopping Mall Intercept Sampling

This is a non-probability sampling method. In this method the respondents are recruited for individual interviews at fixed locations in shopping malls. (For example: Shopper 's Shoppe, Food World, Sunday to Monday). This type of study would include several malls, each serving different socio-economic population.

Example: The researcher may wish to compare the responses of two or more TV commercials for two or more products. Mall samples can be informative for this kind of studies. Mall samples should not be used under following circumstances i.e., if the difference in effectiveness of two commercials varies with the frequency of mall shopping, change in the demographic characteristic of mall shoppers, or any other characteristic. The success of this method depends on "How well the sample is chosen".

Merits

1. It has a relatively small universe.
2. In most cases, it is expected to give quick results. The purpose of deliberate sampling has become a practical method in dealing with economic or practical problems.
3. In studies, where the level of accuracy can vary from the prescribed norms, this method can be used.

Demerits

1. Fundamentally, this is not considered a scientific approach, as it allows for bias.
2. The investigator may start with a preconceived idea and draw samples such that the units selected will be subjected to specific judgment of the enumerator.

Notes

Sequential Sampling

This is a method in which the sample is formed on the basis of a series of successive decisions. They aim at answering the research question on the basis of accumulated evidence. Sometimes, a researcher may want to take a modest sample and look at the results. Thereafter, s(he) will decide if more information is required for which larger samples are considered. If the evidence is not conclusive after a small sample, more samples are required. If the position is still inconclusive, still larger samples are taken. At each stage, a decision is made about whether more information should be collected or the evidence is now sufficient to permit a conclusion.

Example: Assume that a product needs to be evaluated.

A small probability sample is taken from among the current user. Suppose it is found that average annual usage is between 200 to 300 units. It is known that the product is economically viable only if the average consumption is 400 units. This information is sufficient to take a decision to drop the product. On the other hand, if the initial sample shows a consumption level of 450 to 600 units, additional samples are needed for further study.

Quota Sampling

Quota sampling is quite frequently used in marketing research. It involves the fixation of certain quotas, which are to be fulfilled by the interviewers.

Suppose, 2,00,000 students are appearing for a competitive examination. We need to select 1% of them based on quota sampling. The classification of quota may be as follows:

Example: **Classification of Samples**

Category	Quota
General merit	1,000
Sport	600
NRI	100
SC/ST	300
TOTAL	2,000

Quota sampling involves the following steps:

1. The population is divided into segments on the basis of certain characteristics. Here, the segments are termed as cells.
2. A quota of unit is selected from each cell.

Advantages

1. Quota sampling does not require prior knowledge about the cell to which each population unit belongs. Therefore, this sampling has a distinct advantage over stratified random sampling, where every population unit must be placed in the appropriate stratum before the actual sample selection.
2. It is simple to administer. Sampling can be done very quickly.
3. The necessity of the researcher going to various geographical locations is avoided and thus cost is reduced.

Notes

Limitations

1. It may not be possible to get a "representative" sample within the quota as the selection depends entirely on the mood and convenience of the interviewer.
2. Since too much liberty is being allowed to the interviewer, the quality of work suffers if they are not competent.

Snowball Sampling

This is a non-probability sampling. In this method, the initial groups of respondents are selected randomly. Subsequent respondents are being selected based on the opinion or referrals provided by the initial respondents. Further referrals will lead to more referrals, thus leading to a snowball sampling. The referrals will have demographic and psychographic characteristics that are relatively similar to the person referring them.

Example: College students bring in more students on the consumption of Pepsi. The major advantage of snowball sampling is that it monitors the desired characteristics in the population.

Panel Samples

Panel samples are frequently used in marketing research. To give an example, suppose that one is interested in knowing the change in the consumption pattern of households. Samples of households are drawn. These households are contacted to gather information on the pattern of consumption. Subsequently, say after a period of six months, the same households are approached once again and the necessary information on their consumption is collected.

Self Assessment

Fill in the blanks:

7. ..... is also called as the random sampling with replacement.
8. ..... is also called random sampling without replacement.
9. Stratified sampling can be carried out with ..... proportion across the strata proportionate stratified sample.
10. In cluster sampling, ... the units in the selected cluster are studied.

8.4 Distinction between Probability Sample and Non-probability Sample

Probability Sample

1. Here, each member of a universe has a known chance of being selected and included in the sample.
2. Any personal bias is avoided. The researcher cannot exercise his discretion in the selection of sample items.

Example: Random Sample, cluster sample.

Notes

Non-probability Sample

In this case, the likelihood of choosing a particular universe element is unknown. The sample chosen in this method is based on aspects like convenience, quota etc.

Example: Quota sampling, Judgment sampling.

Task Identify the appropriate target population and sampling frame for various situations listed below:

1. The regional marketing manager of a beverage company wants to test market three new flavours to gauge their acceptance.
2. A manufacturer wants to assess whether adequate inventories of spare parts are being maintained by the distributors to prevent shortages and loss of business.
3. A wholesaler dealing with audio/video equipments wants to evaluate the reaction of dealers towards a new promotion policy announced.
4. A TV channel wants to determine the viewing habits of housewives and their programme preferences.
5. A departmental chain such as Food World wants to determine the shopping behaviour of customers who use the credit cards.

8.5 Errors in Sampling

8.5.1 Sampling Error

The only way to guarantee the minimization of sampling error is to choose the appropriate sample size. As the sample keeps on increasing, the sampling error decreases. Sampling error is the gap between the sample mean and population mean.

Example: If a study is done amongst Maruti car-owners in a city to find the average monthly expenditure on the maintenance of car, it can be done by including all Maruti car-owners. It can also be done by choosing a sample without covering the entire population. There will be a difference between the two methods with regard to monthly expenditure.

8.5.2 Non-sampling Error

One way of distinguishing between the sampling and the non-sampling error is that, while sampling error relates to random variations which can be found out in the form of standard error, non-sampling error occurs in some systematic way which is difficult to estimate.

8.5.3 Sampling Frame Error

A sampling frame is a specific list of population units, from which the sample for a study being chosen.

Notes

Example:

- An MNC bank wants to pick up a sample among the credit card holders. They can readily get a complete list of credit card holders, which forms their data bank. From this frame, the desired individuals can be chosen. In this example, sample frame is identical to ideal population namely all credit card holders. There is no sampling error in this case.
- Assume that a bank wants to contact the people belonging to a particular profession over phone (doctors, lawyers) to market a home loan product. The sampling frame in this case is the telephone directory. This sampling frame may pose several problems: (1) People might have migrated. (2) Numbers have changed. (3) Many numbers were not yet listed. The question is "Are the residents who are included in the directory likely to differ from those who are not included"? The answer is yes. Thus in this case, there will be a sampling error.

8.5.4 Non-response Error

This occurs, because the planned sample and final sample vary significantly.

Example: Marketers want to know about the television viewing habits across the country. They choose 500 households and mail the questionnaire. Assume that only 200 respondents reply. This does not show a non-response error, which depends upon the discrepancy. If those 200 who replied did not differ from the chosen 500, there is no non-response error.

Consider an alternative. The people who responded are those who had plenty of leisure time. Therefore, it is implied that non-respondents do not have adequate leisure time. In this case, the final sample and the planned sample differ. If it was assumed that all the 500 chosen have leisure time, but in the final analysis only 200 have leisure time and not others. Therefore, a sample with respect to leisure time leads to response error.

8.5.5 Data Error

This occurs during the data collection, analysis of data or interpretation. Respondents sometimes give distorted answers unintentionally for questions which are difficult, or if the question is exceptionally long and the respondent may not have answer. Data errors can also occur depending on the physical and social characteristics of the interviewer and the respondent. Things such as the tone and voice can affect the responses. Therefore, we can say that the characteristics of the interviewer can also result in data error. Also, cheating on the part of the interviewer leads to data error. Data errors can also occur when answers to open-ended questions are being improperly recorded.

8.5.6 Failure of the Interviewer to Follow Instructions

The respondent must be briefed before beginning the interview, "What is expected"? "To what extent he should answer"? Also, the interviewer must make sure that respondent is familiar with the subject. If these are not made clear by the interviewer, errors will occur.

Editing mistakes made by the editors in transferring the data from questionnaire to computers are other causes for errors.

The respondent could terminate his/her participation in data gathering, because it may be felt that the questionnaire is too long and tedious.

Notes

Did u know? **How to reduce non-sampling error?**

1. For non-response – provide incentives such as a gift or cash. This enhances the possibility as well as incidence of response.
2. Data error: Don't ask question, which respondents cannot answer. Also, do not ask sensitive questions.
3. Train the interviewer to establish a good rapport with the respondents.
4. Avoid leading questions.
5. Pre-test the questionnaire.
6. Modify the sampling frame to make it a representative of the population.

Self Assessment

Fill in the blanks:

11. A ……. is a specific list of population units, from which the sample for a study being chosen.
12. …….. occurs during the data collection, analysis of data or interpretation.

8.6 Sample Size Decision

1. The first factor that must be considered in estimating sample size, is the error permissible.
2. Greater the desired precision, larger will be the sample size.
3. Higher the confidence level in the estimate, the larger the sample must be. There is a tradeoff between the degree of confidence and the degree of precision with a sample of fixed size.
4. The greater the number of sub-groups of interest within the sample, the greater its size must be.
5. Cost is a factor that determines the size of the sample.
6. The issue of response rate: The issue to be considered in deciding the necessary sample size is the actual number of questionnaires that must be sent out. Calculation-wise, we may send questionnaires to the required number of people, but we may not receive the response.

Example: We may like to obtain the family income level from a mail survey, but the researcher may not receive response from everyone. If the researcher feels the response rate is 40%, then he needs to despatch that many extra questionnaires. A low percentage of response can cause serious problems to the researcher. This is known as the non-response error.

Non-response error may be due to 1) failure to locate, 2) flat refusal

The failure to locate: People move to new destinations. However, if the sample frames used are of recent origin, this problem can be overcome.

Flat refusal: We do not know if those who did not respond hold different views or opinions from those who responded.

Notes

This implies that those who don't respond should be motivated. It can be done in any one of the following ways:

1. An advance letter informing the respondents that they will receive a questionnaire and requesting their cooperation. This will generally increase the rate of response.
2. Monetary incentive or gift given to respondents will yield a larger response rate.
3. Proper follow up is necessary after the potential respondent received the questionnaire.

Example: Determine the sample size if standard deviation of the population is 3.9, population mean is 36 and sample mean is 33 and the desired degree of precision is 99%.

Solution: Given $\sigma = 3.9$, $\mu = 36, \bar{x} = 33$ and z = 1% (99% precision implies 1% level of significance)

i.e. z = 2.576 (at 1% l.o.s) (Table value)

We know that sample size n can be obtained using the relation

$$n = \left(\frac{z_\alpha \sigma}{d}\right)^2 \text{ where } d = \mu - \bar{x}$$

$$= \left(\frac{2.576 \times 3.9}{36 - 33}\right)^2 = 11.21 \simeq 11$$

Task Prepare a sample plan including the sample size for Santoor soap, keeping in mind both the male and female customers. Use three economic strata, the educational level and the age group influencing the buyer behaviour. Prepare a sampling design for the following:

1. To measure the effectiveness for a TV Ad on soaps.
2. Consumer reaction to a new brand of coffee introduced.
3. To assess the market share of branded soap.

8.7 Sampling Distribution

A sampling distribution is the probability distribution of a given statistic based on a random sample of certain size n. It may be considered as the distribution of the statistic for all possible samples of a given size. The sampling distribution depends on the underlying distribution of the population, the statistic being considered, and the sample size used. The sampling distribution is frequently opposed to the asymptotic distribution, which corresponds to the limit case .

Example: Consider a normal population with mean and variance. Assume we repeatedly take samples of a given size from this population and calculate the arithmetic mean for each sample – this statistic is called the sample mean. Each sample will have its own average value, and the distribution of these averages will be called the "sampling distribution of the sample mean". This distribution will be normal *N*(m, s2/n) since the underlying population is normal.

The standard deviation of the sampling distribution of the statistic is referred to as the standard error of that quantity. For the case where the statistic is the sample mean, the standard error is:

Notes

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

where σ is the standard deviation of the population distribution of that quantity and n is the size (number of items) in the sample.

Caselet

Case of Food Manufacturer

A leading, food manufacturer inserted a full-page advertisement in a Metro, offering a double coupon deal i.e. the retail store will redeem the company coupon at twice its face value to customers who shopped in Ahar between 8 a.m. to 9 p.m. the next day. A random sample of 200 consumers on the following day of promotion produced the results as below:

a. 100 of the 200 respondents had been to Ahar between 8 a.m. to 9 p.m.

b. Of these 100 respondents, 85 had seen the company's full-page advertisement.

c. Of the remaining 100 respondents, who had not been to Ahar, only 50 had seen the company's ad.

On the basis of the above findings of the survey, the company claimed that, the promotion was a "big success". Do you agree? If so, explain why do you think so?

Self Assessment

Fill in the blanks:

13. Non-response error may be due to 1)……., 2) flat refusal.

14. Sampling distribution is the probability distribution of a given statistic based on a ……of certain size n.

8.8 Summary

- Sample is a representative of population.
- Census represents cent percent of population.
- The most important factors distinguishing whether to choose sample or census is cost and time. There are seven steps involved in selecting the sample.
- There are two types of sample, namely, Probability sampling and Non-probability sample.
- Probability sampling includes random sampling, stratified random sampling systematic sampling, cluster sampling, Multi-stage sampling.
- Random sampling can be chosen by Lottery method or using random number table.
- Samples can be chosen either with equal probability or varying probability.
- Random sampling can be systematic or stratified.
- In systematic random sampling, only the first number is randomly selected.
- Then by adding a constant "K" remaining numbers are generated.

Notes

- In stratified sampling, random samples are drawn from several strata, which has more or less same characteristics.
- In multi-stage sampling, sampling is drawn in several stages.

8.9 Keywords

Census: It refers to complete inclusion of all elements in the population. A sample is a sub-group of the population.

Deliberate Sampling: The investigator uses his discretion in selecting sample observations from the universe. As a result, there is an element of bias in the selection.

Multistage Sampling: The name implies that sampling is done in several stages.

Quota Sampling: Quota sampling is quite frequently used in marketing research. It involves the fixation of certain quotas, which are to be fulfilled by the interviewers.

Random Sampling: Simple random sample is a process in which every item of the population has an equal probability of being chosen.

Sample Frame: Sampling frame is the list of elements from which the sample is actually drawn.

Stratified Random Sampling: A probability sampling procedure in which simple random sub-samples are drawn from within different strata, that are, more or less equal on some characteristics.

8.10 Review Questions

1. What do you analyse as the advantages and disadvantages of probability sampling?
2. Which method of sampling would you use in studies, where the level of accuracy can vary from the prescribed norms and why?
3. Shopping Mall Intercept Sampling is not considered a scientific approach. Why?
4. Quota sampling does not require prior knowledge about the cell to which each population unit belongs. Does this attribute serve as an advantage or disadvantage for Quota Sampling?
5. What suggestions would you give to reduce non sampling error?
6. One mobile phone user is asked to recruit another mobile phone user. What sampling method is this known as and why?
7. Sampling is a part of the population. True/False? Why/why not?
8. Determine the sample size if the standard deviation of population is 20 and the standard error is 4.1.
9. What do see as the reason behind purposive sampling being known as judgement sampling?
10. Suppose, the population consists of 45,000 households, divided into five (5) strata on the basis of monthly income. This can be illustrating as below:

0	-	1000
1001	-	5000
5001	-	7500
7501	-	10,000
		Above 10,000

Notes

Then

(a) Find out the number of units from each strata if the sample constitutes 1% of the population.

(b) If selection is for 150 items selecting equally from each strata, find out the number of sample units from each strata.

Answers: Self Assessment

1.	Sampling	2.	Target
3.	Frame	4.	Small
5.	Target	6.	Resources
7.	Equal Probability	8.	Varying Probability
9.	Same	10.	All
11.	Sampling frame	12.	Data Error
13.	Failure to locate	14.	Random Sample

8.11 Further Readings

Books

Alan T Shao, *Marketing Research*, Cengage.

Cisnal Peter, *Marketing Research*, MCGE.

Cooper and Schinder, *Business Research Methods*, TMH.

CR Kotari, *Research Methodology*, Vishwa Prakashan.

David Luck and Ronald Rubin, *Marketing Research*, PHI.

Goode Hatt, *Methods in Social Science*, McGraw-Hill.

Gupta, *Marketing Research*, ICFAI.

Naresh Amphora, *Marketing Research*, Pearson Education.

S.N. Murthy & U. Bhojanna, *Business Research Methods*, 2nd Edition, Excel Books.

William MC Trochim, *Research Methods*, Biztantra.

William Zikmund, *Business Research Methods*, Thomson.

Online links

www.indiastudychannel.com

www.scribd.com/doc

www.soas.ac.uk

www.web-source.net

Unit 9: Attitude Measurement and Scaling Techniques

Notes

CONTENTS

Objectives
Introduction
9.1 Components of Attitude
9.2 Scaling Technique
9.2.1 Types of Scaling Techniques
9.2.2 Comparative and Non-comparative Scales
9.3 Criteria for the Good Test
9.4 Data Processing Operations
9.4.1 Steps in Processing of Data
9.5 Summary
9.6 Keywords
9.7 Review Questions
9.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Know Basic scaling techniques
- Discuss Comparative and non-comparative scales
- Explain Multi-dimensional scaling techniques
- Identify Limitations of multi-dimensional scaling techniques
- Discuss Criteria for a good test
- Explain Data processing operations

Introduction

Attitude is a degree of positive or negative effect associated with some psychological object. Attitudes are subjective and personal. Attitude influences the behaviour. Purchase decisions are based upon the attitudes. The attitudes can change over time.

9.1 Components of Attitude

Attitude has three components, namely cognitive, affective and the behavioural.

- ***Cognitive:*** This refers to the respondents' beliefs, knowledge or awareness about an event or an object. This is usually acquired from friends, periodicals etc. Sometimes, it is also known as the belief component. Statements like: (a) I am aware of the product 'X' (b) I have no idea about the product 'B' (c) That institute is excellent.

  Cognitive component is very important in marketing in terms of creating awareness about the product, person, etc.

Notes

- ***Affective:*** This refers to the respondent's liking or preferences for an object. This is also known as the feeling component. (a) I like the product 'A' (b) Advertisement 'X' is poor. This component reveals the buyers' positive or negative attitude towards the product.
- ***Behavioural:*** This refers to the respondent's intention to buy. This is a situation prior to the purchase. In marketing, the usage and buying pattern depends on this component. This is also known as action component.

Did u know? **What are the Determinants of Attitude? (What Alters the Attitude?)**

Attitudes are not static, but change continuously. Attitudes undergo change due to five factors:

- Information gathered in the past relating to the actual experience
- Individual perception and belief
- Exposure to new information
- Changes in the group membership
- Individual personality.

9.2 Scaling Technique

The generation of a continuum upon which measured objects are located.

- A quantifying measure – a combination of items that is progressively arranged according to value or magnitude.
- Purpose is to quantitatively represent an item's, person's, or event's place in the scaling continuum.

9.2.1 Types of Scaling Techniques

These are four kinds of scales, namely:

(a) Nominal scale

(b) Ordinal scale

(c) Interval scale

(d) Ratio scale

Nominal Scale

In this scale, numbers are used to identify the objects. For example, University Registration numbers assigned to students, numbers on their jerseys.

Example: Have you ever visited Bangalore?

Yes-1

No-2

'Yes' is coded as 'One' and 'No' is coded as 'Two'. The numeric attached to the answers has no meaning, and is a mere identification. If numbers are interchanged as one for 'No' and two for

Notes

'Yes', it won't affect the answers given by respondents. The numbers used in nominal scales serve only the purpose of counting.

The telephone numbers are an example of nominal scale, where one number is assigned to one subscriber. The idea of using nominal scale is to make sure that no two persons or objects receive the same number. Similarly, bus route numbers are the example of nominal scale.

"How old are you"? This is an example of a nominal scale.

"What is your PAN Card number?

Arranging the books in the library, subject wise, author wise – we use nominal scale.

Example: Physics-48, Chemistry-92, etc.

Caution It should be kept in mind that nominal scale has certain limitation, viz.

(a) There is no rank ordering.

(b) No mathematical operation is possible.

(c) Statistical implication – Calculation of the standard deviation and the mean is not possible. It is possible to express the mode.

Ordinal Scale (Ranking Scale)

The Ordinal scale is used for ranking in most market research studies. Ordinal scales are used to ascertain the consumer perceptions, preferences, etc. For example, the respondents may be given a list of brands which may be suitable and were asked to rank on the basis of ordinal scale.

- Lux
- Liril
- Cinthol
- Lifebuoy
- Hamam

Rank	Item	Number of Respondents
I	Cinthol	150
II	Liril	300
III	Hamam	250
IV	Lux	200
V	Lifebuoy	100
Total		**1,000**

In the above example, II is mode and III is median.

Statistical implications: It is possible to calculate the mode and the median.

In market research, we often ask the respondents to rank the items, like for example, "A soft drink, based upon flavour or colour". In such a case, the ordinal scale is used. Ordinal scale is a ranking scale.

Notes

Rank the following attributes of 1–5 scale according to the importance in the microwave oven:

Attributes	Rank
(a) Company Image	5
(b) Functions	3
(c) Price	2
(d) Comfort	1
(e) Design	4

Did u know? ***What is the difference between nominal and ordinal scales?***

In nominal scale numbers can be interchanged, because it serves only for the purpose of counting. Numbers in Ordinal scale have meaning and it won't allow interchangeability.

Interval Scale

Interval scale is more powerful than the nominal and ordinal scales. The distance given on the scale represents equal distance on the property being measured. Interval scale may tell us "How far the objects are apart with respect to an attribute?" This means that the difference can be compared. The difference between "1" and "2" is equal to the difference between "2" and "3".

Example:

1. Suppose we want to measure the rating of a refrigerator using interval scale. It will appear as follows:

1.	Brand name	Poor ————————	Good
2.	Price	High ————————	Low
3.	Service after-sales	Poor ————————	Good
4.	Utility	Poor ————————	Good

   The researcher cannot conclude that the respondent who gives a rating of 6 is 3 times more favourable towards a product under study than another respondent who awards the rating of 2.

2. How many hours you spend to do class assignment every day?

   < 30 min.

   30 min. to 1 hr.

   1hr. to 1½ hrs.

   > 1½ hrs.

Statistical implications: We can compute the range, mean, median, etc.

Task Analyse the difference between interval and ordinal scales.

Ratio Scale

Ratio scale is a special kind of internal scale that has a meaningful zero point. With this scale, length, weight or distance can be measured. In this scale, it is possible to say, how many times greater or smaller one object is being compared to the other.

Notes

Example: Sales this year for product A are twice the sales of the same product last year.

Statistical implications: All statistical operations can be performed on this scale.

9.2.2 Comparative and Non-comparative Scales

In non-comparative scale, the respondent is allowed to arbitrarily apply different standards. In other words, different reference points are chosen. This may lead to ambiguities in computation.

To overcome this problem, in comparative scale, a reference point is fixed to facilitate comparison.

Illustration of comparative and non-comparative scale is shown below.

Comparative Scale: In each of the following, which store do you think is better (please tick one store from the following)

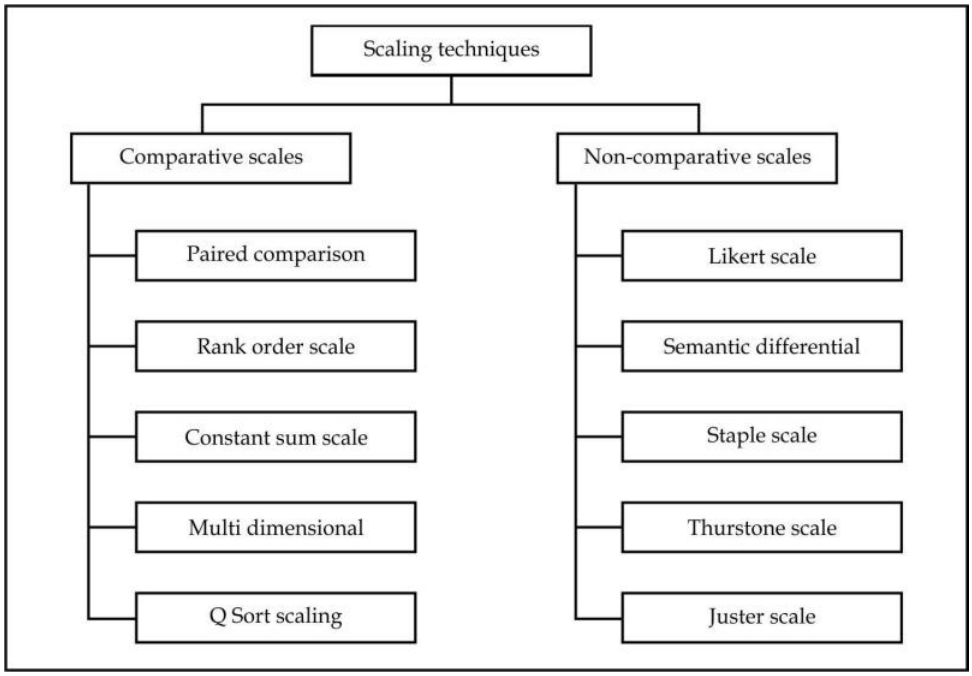
(a) Food world……. (b) Spencer……

(c) Total …….. (d) Fresh……

(e) Smart……. (f) More…….

Non-comparative Scale: The most important reason for shopping at Big Bazar is ….

(a) Ambiance (b) Price (c) Variety (d) Parking space (e) Discount (f) Home delivery



Comparative Scales

Paired Comparison

Example: Here a respondent is asked to show his preferences from among five brands of coffee – A, B, C, D and E with respect to flavors. He is required to indicate his preference in pairs. A number of pairs are calculated as follows. The brands to be rated are presented two at a time,

Notes

so each brand in the category is compared once to every other brand. In each pair, the respondents were asked to divide 100 points on the basis of how much they liked one compared to the other. The score is totally for each brand.

$$\text{No. of pairs} = \frac{N(N-1)}{2}$$

$$\text{In this case, it is} = \frac{5(5-1)}{2} = 2$$

A&B	B&D
A&C	B&E
A&D	C&D
A&E	C&E
B&C	D&E

If there are 15 brands to be evaluated, then we have 105 paired comparison(s) and that is the limitation of this method.

Rank Order Scale

In this method, respondents are required to rank more than two objects or alternatives based on some criteria.

Example: Good taste, ease of usage, etc.

It is simpler than paired comparison scale, as its procedure can be easily understood by the respondent.

Rank order scales are more difficult than rating scale because they involve comparison and hence require more attention and mental effort. The main disadvantage is the respondent may not like to make a choice among the given alternative and hence compelled to choose one of the given alternative.

Another shortcoming of the rank order scaling is that, respondents cannot meaningfully rank more than 5 to 6 objects. The problem will not arise while ranking of first and last objects but with those in the undifferentiated middle. When there are several objects, one solution is to divide the ranking into 2 stages. For example, with 9 objects, the first stage would be to rank the objects into classes. Top three middle three and last three. The next stage would be to rank the 3 objects within each class.

Constant Sum Scale

Constant sum scale is one of the methods of comparative scaling. In this method, the respondent is instructed to allocate some constant sum (points) to various features given, based on the importance of attribute to the respondent. For example, number of features in selection of a bank by the respondent may be done as follows. 100 points are assigned, which will be allocated among the features. The features may be as follows:

	Feature	No. of points (sum)
1.	Location	– – – – –
2.	Banking hours	– – – – –
3.	Interest rate	– – – – –

Contd...

Notes

4.	Courteous nature of employees	_ _ _ _ _
5.	Loan facilities	_ _ _ _ _
		_ _ _ _ _
		100

Banking features as given above are allocated 100 points. Allocation depends on the importance of features as judged by the respondent. Respondent may allocate interest rate 30 points, location 10 points, etc. Here respondent need to divide 100 points among various features given above.

By using his scale, bank will come to know the attributes that are more important to the customer which in turn influence the customer to choose a particular bank. The only precaution to be taken while administering this scale is that, if there are too many attributes, the respondent will find it difficult, since lot of mental energy is required to answer the scale. This scale can not be used effectively in case the respondent is illiterate.

Multi-dimensional Scaling

This is used to study consumer attitudes, particularly with respect to perceptions and preferences. These techniques help identify the product attributes that are important to the customers and to measure their relative importance. Multi-Dimensional Scaling is useful in studying the following:

1. (a) What are the major attributes considered while choosing a product (soft drinks, modes of transportation)? (b) Which attributes do customers compare to evaluate different brands of the product? Is it price, quality, availability, etc.?
2. Which is the ideal combination of attributes according to the customer? (i.e., which two or more attributes consumer will consider before deciding to buy).
3. Which advertising messages is compatible with the consumer's brand perceptions?

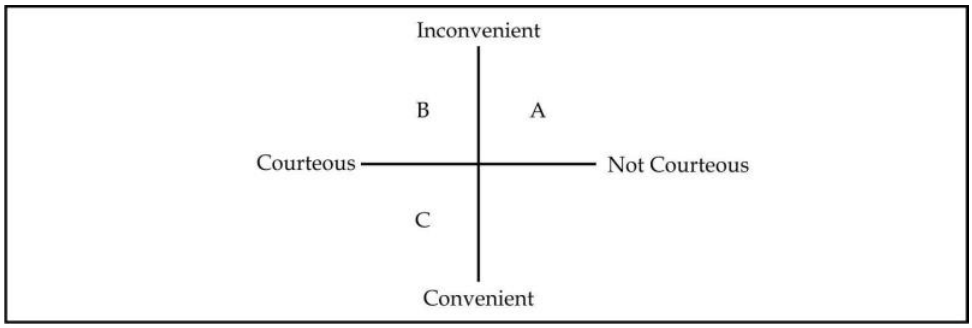
This scaling is used to describe similarity and preference of brands. The respondents were asked to indicate their perception, or the similarity between various objects (products, brands, etc.) and preference among objects. This scaling is also known as perceptual mapping.

There are two ways of collecting the input data to plot perceptual mapping:

1. Non-attribute method.
2. Attribute method.

1. ***Non-attribute method:*** Here, the researcher asks the respondent to make a judgment about the objects directly. In this method, the criteria for comparing the objects is decided by the respondent himself.
2. ***Attribute method:*** In this method, instead of respondents selecting the criteria, they were asked to compare the objects based on the criteria specified by the researcher.

For example, to determine the perception of a consumer: Assume there are five insurance companies to be evaluated on two attributes namely (1) convenient locality (2) courteous personal service. Customers' perception regarding the five insurance companies are as follows:



Notes

A, B, C, D and E are five insurance companies.

According to the map, B & E are dissimilar insurance companies.

C is being located very conveniently.

A is a less convenient in location compared to E.

D is a less convenient in location than C.

E is a less convenient location compared to D.

Did u know? **What tools are used in MDS?**

Software such as SPSS, SAS and Excel are the packages used in MDS. Brand positioning research is one of SPSS's important features. SAS is a business intelligence software. Excel is also used to a certain extent.

Caselet

Case: New Baby Care Product (Perceptual Mapping)

This method is particular about the steps adopted by searchers to assist a company in the newborn baby care market. The example cited is that of Marico. This map helps Marico to identify the position of its competitors. Marico introduced a new brand baby oil named "Sparsh." This is an unorthodox entry. Marico was the first to rope in an ambassador actress in a market worth ₹ 300 crore. A second brand ambassador to speak in favour was Sonali Bendre, for both baby oil and a bathing bar. The reason for choosing a female ambassador was to lay emphasis on the concept of motherhood.

Marico is a leader in the world's largest coconut oil brand namely, Parachute. They are now switching over to health care products from hair oil and edible oil. Though adult health care products constitute a ₹. 1,500 crore market, baby care segment still continues to be a niche market. The following are some of the obstacles in developing loyalty towards the baby care products.

1. The family may use the same adult hair care product for children as well.
2. Customers repeating the product are hard to find due to the fact that women have fewer babies in present times.
3. Women stick to the product that their mothers recommend.
4. Herbal versions are still popular in urban, semi-urban and rural areas.
5. There are big players in the field of baby care products.

A few example are Johnson and Johnson, Dabur, Wipro, etc. The market also has the Himalaya Drug Company which has established its own herbal baby care division.

The uniqueness of Sparsh lies in the fact that it meets consumer needs by using traditional ingredients in modern packing. Marico's main effort is to create brand differentiation.

Two parameters used by Sparsh of Marico are:

1. Price.
2. Value perception.

Contd...

Notes

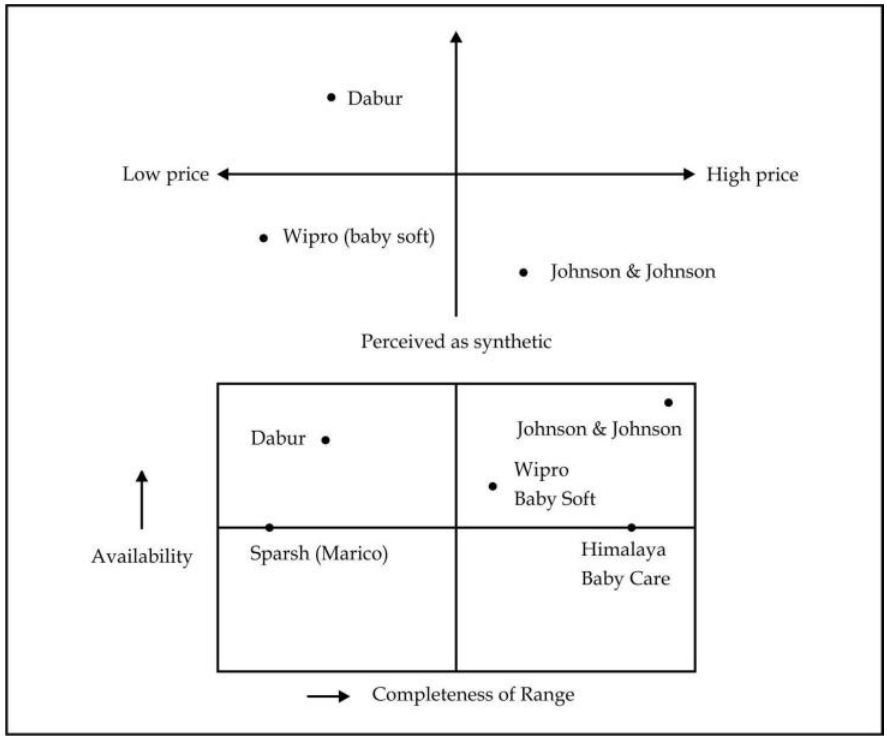
In a segment, where perceived quality governs decision-making, and value as a parameter is the choice, value proposition is central to Sparsh's marketing.

In a segment, where price is the consideration, the company has priced it on par with leaders.

In promotion, the company has used a two-pronged approach:

1. Build brand value.
2. Cut through the clutter.

Perceptual Mapping for Baby care Brands in India



Source: Times of India

Use of Multi-dimensional Scaling:

1. To determine salient product attributes perceived by buyers in the market.
2. To know the combination of attributes buyers are likely to prefer.
3. To understand the products which are viewed as substitutes and those that are differentiated.
4. For segmenting the market.

Limitations of MDS:

1. ***Conceptual problem:*** The criteria on which the similarities are gauged may vary during an interview with respondents. They vary depending on what the respondent thinks. A customer may buy something for himself or he may gift a product to others. In both cases, the criteria used for selection are different.
2. ***Preference:*** Keeps changing from time to time.
3. Complicated computational problem.

Notes

Q Sort Scaling

When there are very large number of characteristics to be rated, it becomes very difficult for the respondent to rank order. To deal with this, Q-Sort scaling procedure is used. In this technique, respondents are used to sort out the various characteristics into convenient groups. Therefore, large number of groups is used in this method. This will increase the reliability of the results.

Suppose respondents are given say 100 statements. They are asked to place them in eleven piles, ranging from "most highly agree" to "least highly agree".

Caution The ideal number of this type of scaling should be between sixty and ninety.

The number of statements/objects placed in each pile is pre specified so that roughly normal distribution of object is obtained.

Non-comparative Scales

Likert Scale

It is known as summated rating scale. This consists of a series of statements concerning an attitude object. Each statement has '5 points', Agree and Disagree on the scale. They are also called summated scales, because scores of individual items are summated to produce a total score for the respondent. The Likert Scale consists of two parts – item part and evaluation part. Item part is usually a statement about a certain product, event or attitude. Evaluation part is a list of responses like "strongly agree" to "strongly disagree". The five point-scale is used here. The numbers like +2, +1, 0, –1, –2 are used. The Likert Scale must contain an equal number of favourable and unfavourable statements, Now, let us see with an example how the attitude of a customer is measured with respect to a shopping mall.

Table 9.1 Evaluation of Globus – The Super Market by Respondent

Sl. No.	Likert scale items	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1	Salesmen at the shopping mall are courteous	-	-	-	-	-
2	Shopping mall does not have enough parking space	-	-	-	-	-
3	Prices of items are reasonable.	-	-	-	-	-
4	Mall has wide range of products to choose	-	-	-	-	-
5	Mall operating hours are inconvenient	-	-	-	-	-
6	The arrangement of items in the mall is confusing	-	-	-	-	-

The respondents' overall attitude is measured by summing up his (her) numerical rating on the statement making up the scale. Since some statements are favourable and others unfavourable, it is the one important task to be done before summing up the ratings. In other words, "strongly agree" category attached to favourable statement and "strongly disagree" category attached to

Notes

unfavourable. The statement must always be assigned the same number, such as +2, or –2. The success of the Likert Scale depends on "How well the statements are generated?" The higher the respondent's score, the more favourable is the attitude. For example, if there are two shopping malls, ABC and XYZ and if the scores using the Likert Scale are 30 and 60 respectively, we can conclude that the customers' attitude towards XYZ is more favourable than ABC.

Semantic Differential Scale

This is very similar to the Likert Scale. It also consists of a number of items to be rated by the respondents. The essential difference between Likert and Semantic Differential Scale is as follows:

It uses "Bipolar" adjectives and phrases. There are no statements in the Semantic Differential Scale.

Each pair of adjective is separated by a seven point scale.

Notes Some individuals have favourable descriptions on the right side, while some have on the left side. The reason for the reversal is to have a combination of both favourable and unfavourable statements.

Task Please rate the five real estate developers mentioned below on the given scales for each of the five aspects. Developers are:

(1) Ansal (2) Raheja (3) Purvankara (4) Mantri (5) Salpuria

		–3	–2	–1	0	+1	+2	+3	
1.	Not reliable	_	_	_	_	_	_	_	Reliable
2.	Expensive	_	_	_	_	_	_	_	Not expensive
3.	Trustworthy	_	_	_	_	_	_	_	Not trustworthy
4.	Untimely delivery	_	_	_	_	_	_	_	Timely delivery
5.	Strong Brand Image	_	_	_	_	_	_	_	Poor brand image

The respondents were asked to tick one of the seven categories which describes their views on attitude. Computation is being done exactly the same way as in the Likert Scale. Suppose, we are trying to evaluate the packaging of a particular product. The seven point scale will be as follows:

" I feel ……..

1. Delighted
2. Pleased
3. Mostly satisfied
4. Equally satisfied and dissatisfied
5. Mostly dissatisfied
6. Unhappy
7. Terrible.

Notes

Staple Scale

This scale is a modified version of semantic differential scale. It uses only one pole. It is a ten print scale with a range of +5 to -5. This scale measures both the direction and intensity of attitude simultaneously. Unlike semantic differential study, which uses bipolar adjectives, here single word is used to describe the characteristic of interest. There is no absolute zero point. This is an interval scales. Respondents are asked to indicate the object by selecting a numerical response category. The main advantage of this scale is that it is simple to administer and also to construct.

An illustration of staple scale is as follows. You have been associated with M/s XYZ company, conducting marketing research.

How would you rate M/s XYZ Ltd. in terms of

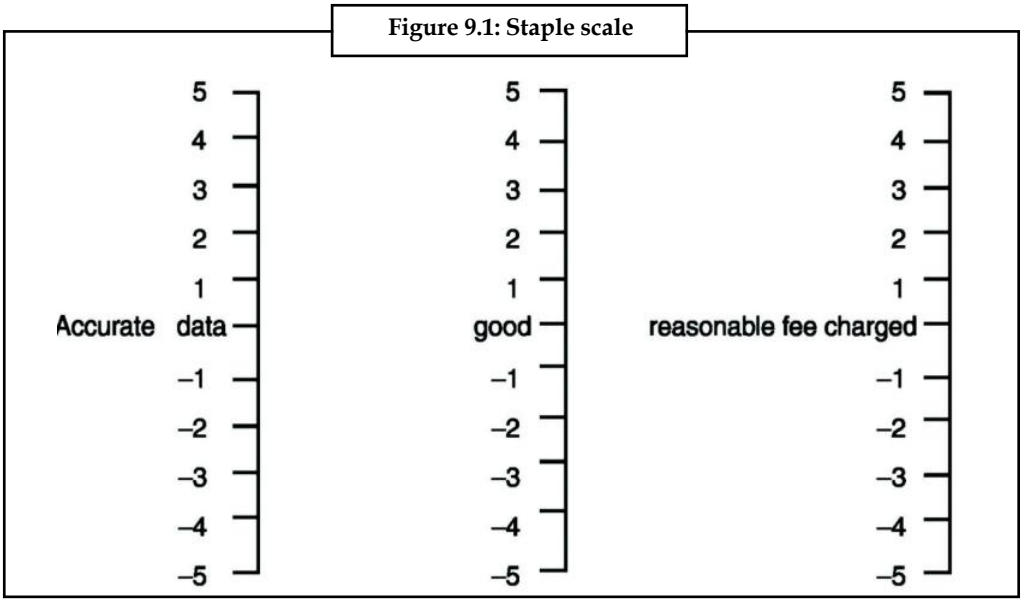
(a) Accuracy of data provided by them

(b) Quality of researchers employed by them

(c) Quantum of money charged for providing data

Circle the number you think is most appropriate. If you think, the data provided by the research company is extremely accurate circle +5 and vice-versa.

Figure 9.1: Staple scale



Staple Scale is used in developing profile analysis. Despite the simplicity of constructing and usage, semantic differential scale finds an edge.

Thurstone Scale

This is also known as an equal appearing interval scale. The following are the steps to construct a Thurstone Scale:

Step 1: To generate a large number of statements, relating to the attitude to be measured.

Step 2: These statements (75 to 100) are given to a group of judges, say 20 to 30, who were asked to classify them according to the degree of favourableness and unfavourableness.

Step 3: 11 piles are to be made by the judges. The piles vary from "most unfavourable" in pile 1 to neutral in pile 6 and most favourable statement in pile 11.

Notes

Step 4: Study the frequency distribution of ratings for each statement and eliminate those statements, which different judges have given widely scattered ratings.

Step 5: Select one or two statements from each of the 11 piles for the final scale. List the selected statements in random order to form the scale.

Step 6: The respondents whose attitudes are to be scaled were given the list of statements and asked to indicate their agreement or disagreement with each statement. Some may agree with one statement while some may agree with more than one statement.

Example:

1. Crime and violence in movies:
    1. All movies with crime and violence should be prohibited by law.
    2. Watching crime and violence in movies is a waste of time.
    3. Most movies with crime are bad and harmful.
    4. The direction and theme in most crime movies are monotonous.
    5. Watching a movie with crime and violence does not interfere with my routine life.
    6. I have no opinion one way or the other, about watching movies with crime and violence.
    7. I like to watch movies with crime and violence.
    8. Most movies with crime and violence are interesting and absorbing.
    9. Crime movies act as a knowledge bank gained by the audience.
    10. People learn "how to be safe and protect oneself" by seeing a movie on crime.
    11. Watching crime in a movie does not harm our lifestyle.

    Conclusion: A respondent might agree with statements 8, 9 and 10. Such agreement represents a favourable attitude towards crime and violence. On the contrary, if items 1, 3, 4 are chosen by respondents, it shows that respondents are unfavourably disposed towards crime in movies. If the respondent chooses 1, 5 and 11, it could be interpreted to indicate that she (he) is not consistent in his(her) attitude about the subject.

2. Suppose, we are interested in the attitude of certain socio-economic class of respondents towards savings and investments. The final list of statements would be as follows:
    1. One should live for the present and not the future. So, savings are absolutely not required.
    2. There are many attractions to spend the money saved.
    3. It is better to spend savings than risk them in investments.
    4. Investments are unsafe as the money is also blocked.
    5. You earn to spend and not to invest.
    6. It is not possible to save these days.

Notes

7. A certain amount of income should be saved and invested.
8. The future is uncertain and investments will protect us.
9. Some amount of savings and investments are a must for every individual.
10. One should try to save more so that most of it could be invested.
11. All savings should be invested for the future.

Conclusion: A respondent agreeing to statements 8, 9 and 11 would be considered having a favourable attitude towards savings and investments. The person agreeing with statements 2, 3 and 4 is an individual with an unfavourable attitude. Also, if a respondent chooses statements 1, 3, 7 or 9, his attitude is not considered consistent.

Merits of Thurstone Scale:

1. Very reliable, if we are measuring a single attitude.
2. Used to find attitude towards issues like war, religion, language, culture, place of worship, etc.

Limitations:

1. Limited use in marketing research, since it is time consuming.
2. Collecting a number of statements (100-200) makes it a very tedious job.
3. Bias on the part of the judges cannot be avoided.
4. It is an expensive method.

Juster's 11-point Probability Scale (Juster Scale)

In 1966, Professor F. Thomas Juster argued that since verbal intentions are simply disguised probability statements, then why not directly capture the probabilities themselves as measured by the respondents.

Juster's 11 point probability scale can be used to produce estimates of the average probability that a population will do something by a future time. Since what is being measured is a probability, the mean response estimates the proportion of the population that will perform the behaviour at issue.

An example is given by the question, "On a scale of 0-10 where 0 indicates no chance and 10 indicates certainty, what is the chance that you will change your primary bank in the next 12 months?" If then the average response is 3.2, this translates to 32% of the population intend to switch banks.

The Juster scale in its many applications has been found to be superior as a predictive measure of future purchase behaviour than other intentions scales. The distribution of responses, however, has been found to affect the predictive accuracy of the scale. Not surprisingly, the greater the variation in responses, the less accurate the predictions.

Studies have shown that purchase probabilities can be over or under estimated by the Juster scale, but on average, it is the most consistent in accurately predicting actual purchase rates. There are important issues to be considered in the administration of the Juster Scale that have been found to contribute to variation in its effectiveness. These include unfamiliarity of the respondent with new products, training of the administrator and differences in age and education level of respondents.

Notes

The Juster scale has also been successfully used to predict respondent behaviour outside the typical consumption behaviour realm, which includes being applied to telephone surveys, fast moving consumer goods, self-completion questionnaires, services, brands and customer loyalty. One example of such an extension involved predicting the percentage of a given population of adults currently at home looking after children, who will take up paid employment in the next year. At an aggregate level in this example, the Juster Scale mean was 1.9 indicating that a predicted 19% of respondents would find paid work in the next year. When actual behaviour was measured in the following year, it was found that indeed, 19% of these respondents were in paid employment.

Table 9.2: Juster's 11-point Probability Scale

Score	Verbal equivalent
0	No chance, almost no chance (1 in 100)
1	Very slight possibility (1 chance in 10)
2	Slight possibility (2 chances in 10)
3	Some possibility (3 chances in 10)
4	Fair possibility (4 chances in 10)
5	Fairly good possibility (5 chances in 10)
6	Good possibility (6 chances in 10)
7	Probable (7 chances in 10)
8	Very probable (8 chances in 10)
9	Almost sure (9 chances in 10)
10	Certain, practically certain (99 chances in 100)

Task A manufacturer of packed bakery items wants to evaluate customer attitudes toward his product brand. 300 customers who buy this brand filled the questionnaire that was sent to them. The answers of this questionnaire were converted to scale and the results are as follows:

1. The average score from the above sample on a 10-item *Likert Scale* was 65.
2. Average score for a sample on 10-item *Semantic Differential Scale* was 60.

You are required to indicate whether these customers had a favourable or unfavourable attitude towards the product.

Self Assessment

Fill in the blanks:

1. ..... scale may tell us "How far the objects are apart with respect to an attribute?"
2. ..... scale is used to assess attitude of the respondents group regarding any issue of public interest.
3. ..... Scaling is used to study consumer attitudes, particularly with respect to perceptions and preferences.
4. Thurstone Scale is also known as an ..... scale.

Notes

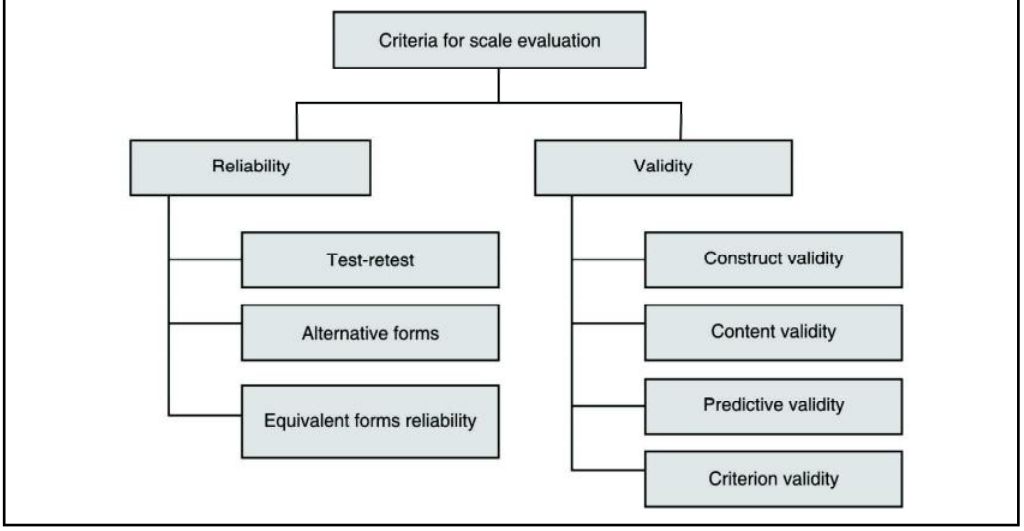
5. Semantic Differential Scale is very similar to the ..... Scale.

6. The Likert Scale consists of two parts – ..... and .....

9.3 Criteria for the Good Test

There are two criteria to decide whether the scale selected is good or not. They are shown in the diagram given below:

Figure 9.2: Criteria for the Good Test



Reliability

For a research to be sound, reliability is an important criterion. Reliability means, ability of a scale to produce a consistent result if repeated measurements are taken. Reliability is the extent to which scales are free from random error and produce consistent results. Lesser the random error, the data will be more reliable. Random error is not a constant error but, it depends on the person using the measuring instrument or measurement situation. A random error leads to inconsistency in measurement; when the measurements are made on the same objects or persons.

An example of random error would be the use of elastic scale to measure a person's height. If 2 successive measurements are made, the person who is measuring would stretch the elastic ruler to different degree on 2 successive occasions, therefore, the height measurement would never be the same, even though the person's height has not changed.

In addition to the random error, there is one more error namely systematic error. It is also called as non sampling error. This error includes all types of errors except random sampling error. Therefore, we can say

Measurement result = true measurement + measurement error

Measurement error = true measurement + systematic error + random error

Test-retest Method

There is an approach called test-retest to check the reliability. In this approach, respondents are given identical sets of scales at two different points of time under almost identical conditions. The time interval is between 2 to 5 weeks. The similarity between 2 measurements is determined by calculating the correlation coefficient. Higher the value of correlation coefficient, greater the reliability.

Notes

The disadvantage of this method is that, if the interval between first and second test is more the scale will be less reliable.

Second disadvantage is, it is difficult to convince the original respondents to take the test for second time.

Third disadvantage is that, the second time answer may be influenced by the first time answer. Assume that, an opinion about the hospital service is asked. Two weeks later, if the same question is asked, the reply by respondent will be influenced by what was told the first time.

Equivalent Forms Reliability

In this method 2 "equivalent" scales are used to obtain consistent results. So, the researcher administer one scale to the respondent and 2 weeks later another scale, which is equivalent of the first one to the same respondent.

The greatest problem of this method is to construct 2 scales that appear to be different but have similar effect. This alternative forms test is similar to the test-retest method, except that the test-retest method uses the same measuring instrument both the times.

Internal Consistency

In this method two or more measurement of the same concept is taken at the same time and then compares to see if they agree with each other. Suppose we use Likert scale and offer choices from strongly agree to strongly disagree to determine consumer attitude towards the service rendered by Big Bazar. Suppose the researcher prepares 4 statements scale to measure this:

1. I enjoy shopping at Big Bazar.
2. All my needs for my household are satisfied by shopping in Big Bazar.
3. Service provided to me by Big Bazar is excellent.
4. I like the front line salesman regarding the service rendered.

The degree to which the 4 statements show correlation across a sample of respondents indicates the reliability of the measure. If correlation is high, then reliability will also be high.

Methods to Improve Reliability

i. Number of measurements is to be increased. Instead of conducting one test, average scores of several equivalent forms of the test is to be considered for reliability.

ii. Controls used for conducting the experiment must be good. Example, (a) measuring device must be accurate. (b) The researcher who administers must be trained to avoid bias in respondents.

iii. Items to be measured must be stated clearly.

Validity

The paradigm of validity focused in the question "Are we measuring, what we think, we are measuring"? Success of the scale lies in measuring "What is intended to be measured?" Of the two attributes of scaling, validity is the most important.

There are several methods to check the validity of the scale used for measurement.

1. ***Construct Validity:*** A sales manager believes that there is a clear relation between job satisfaction for a person and the degree to which a person is an extrovert and the work performance of his sales force. Therefore, those who enjoy high job satisfaction, and have

Notes

extrovert personalities should exhibit high performance. If they do not, then we can question the construct validity of the measure.

2. ***Content Validity:*** A researcher should define the problem clearly. Identify the item to be measured. Evolve a suitable scale for this purpose. Despite these, the scale may be criticised for being lacking in content validity. Content validity is known as face validity. An example can be the introduction of new packaged food. When new packaged food is introduced, the product representing a major change in taste. Thousands of consumers may be asked to taste the new packaged food. Overwhelmingly, people may say that they liked the new flavour. With such a favourable reaction, the product when introduced on a commercial scale may still meet with failure. So, what is wrong? Perhaps a crucial question that was omitted. The people may be asked if liked the new packaged food, to which the majority might have "yes" but the same respondents were not asked, "Are you willing to give up the product which you are consuming currently?" In this case, the problem was not clearly identified and the item to be 'measured' was left out.

3. ***Predictive Validity:*** This pertains to "How best a researcher can guess the future performance from the knowledge of attitude score"?

Example: An opinion questionnaire, which is the basis for forecasting the demand for a product has predictive validity. The procedure for predictive validity is to first measure the attitude and then predict the future behaviour. Finally, this is followed by the measurement of future behaviour at an appropriate time. Compare the two results (past and future). If the two scores are closely associated, then the scale is said to have predictive validity.

4. ***Criterion Validity:***

   (a) Examines whether measurement scale performs as expected in relation to other variables selected as meaningful criteria i.e., predicted and actual behavior should be similar

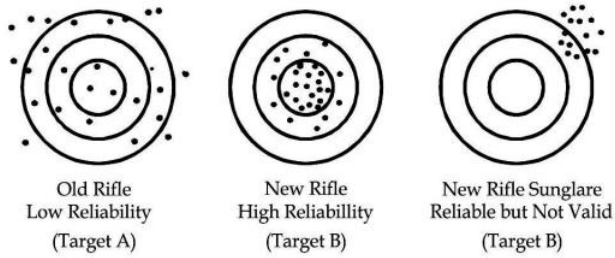
   (b) Addresses the question of what construct or characteristic the scale is actually measuring

5. ***Convergent Validity:*** Extent to which scale correlates positively with other measures of the same construct.

6. ***Discriminant Validity:*** Extent to which a measure does not correlate with other constructs from which it is supposed to differ.

7. ***Nomological Validity:*** Extent to which scale correlates in theoretically predicted ways with measures of different but related constructs.

Figure 9.3: Reliability and Validity on Target



Notes

Did u know? **What is the relationship between reliability and validity?**

1. If a scale is to be valid, it must be reliable.
2. The scale does not have to be valid to be reliable.
3. Reliability is a necessary but not a sufficient condition for validity. This is because validity requires other factors to be satisfied.
4. Validity is not a necessary condition for reliability.

Task Suppose a cosmetic manufacturing company wants to ascertain the perception of its customers towards a product. Take the 7-item scale to measure the perceived perception of the product using Likert and Semantic Differential scales. The following are some of the likely adjectives which are used in Semantic Differential scale:

Unfavourable	**X**	**Favourable**
Soft	X	Hard
Organised	X	Disorganised
Quick	X	Slow
Formal	X	Informal
Pleasure	X	Displeasure
Complex	X	Simple
Cheap	X	Costly
Pleasant	X	Unpleasant
Fragrant	X	Less fragrant
Dominating	X	Submissive
Rational	X	Emotional

The normal categories for Semantic Differential will be as follows:

1. Most strongly agree
2. Strongly agree
3. Agree
4. Undecided
5. Disagree
6. Strongly disagree
7. Most strongly disagree.

Self Assessment

Fill in the blanks:

7. An .....questionnaire, which is the basis for forecasting the demand for a product has predictive validity.

Notes

8. Those who enjoy high job satisfaction, and have extrovert personalities should exhibit .....performance.
9. Reliability deals with .....and.....
10. There are two criteria to decide whether the scale selected is good or not, viz. .....and .....
11. The advantage and disadvantages of a Stapel scale, as well as the results, are very similar to those for a ..... differential.
12. In ……method two or more measurement of the same concept is taken at the same time and then compares to see if they agree with each other.

9.4 Data Processing Operations

Processing data is very important in market research. After collecting the data, the next task of the researcher is to analyse and interpret the data. The purpose of analysis is to draw conclusions. There are two parts in processing the data:

(1) Data analysis

(2) Interpretation of data

Analysis of the data involves organising the data in a particular manner. Interpretation of data is a method for deriving conclusions from the data analysed. Analysis of data is not complete, unless it is interpreted.

9.4.1 Steps in Processing of Data

1. Preparing raw data
2. Editing
3. Coding
4. Tabulation
5. Summarising the data
6. Usage of statistical tools.

Preparing Raw Data

Data collection is a significant part of market research. Even more significant is to filter out the relevant data from the mass of data collected. Data continues to be in raw form, unless they are processed and analysed.

Primary data collected by surveys and observations by field investigations are hastily entered into questionnaires. Due to the pressure of interviewing, the researcher has to write down the responses immediately. Many times this may not be systematic. The information so collected by field staff is called raw data.

The information collected may be illegible, incomplete and inaccurate to a considerable extent. Also the information collected will be scattered in several data collection formats. The data lying in such a crude form are not ready for analysis. Keeping this in mind, the researcher must take some measures to organise the data so that it can be analysed.

The various steps which are required to be taken for this purpose are (a) editing and (b) coding and (c) tabulating.

Editing

Notes

The main purpose of editing is to eliminate errors and confusion. Editing involves inspection and correction of each questionnaire. The main role of editing is to identify commissions, ambiguities and errors in response.

Editing thus means the activity of inspecting, correcting and modifying the correct data.

This can be done in two stages (a) Field editing (b) Office editing.

(a) ***Field editing:*** Objectives of field editing are – To make sure that proper procedure is followed in selecting the respondent, interview them and record their responses. In field editing, speed is the main criteria, since editing should be done when the study is still under progress. The main problems faced in field editing are:

(1) Inappropriate respondents

(2) Incomplete interviews

(3) Improper understanding

(4) Lack of consistency

(5) Legibility

(6) Fictitious interview

Examples:

1. *Inappropriate respondents:* It is intended to include house owners in the sample for conducting the survey. If a tenant is interviewed, it would be wrong.
2. *Incomplete interview:* All questions are to be answered. There should not be any 'blanks'. Blanks can have different meanings, like (a) No answer (b) Refusal to answer (c) Question not applicable (d) Interviewer by oversight did not record. The reason for no answer could be that the respondent really does not know the answers. Sometimes, the respondent does not answer, may be because of the sensitive or emotional aspect of the question.
3. *Lack of understanding:* The interviewer, in a hurry, would have recorded some abbreviated answer. Later at the end of the day, s(he) cannot figure out what it meant.
4. *Consistency:* The earlier part of the questionnaire indicates that there are no children and in the later part the age of children is mentioned.
5. *Legibility:* If what is said is not clear, the interviewer must clarify the same on the spot.
6. *Fictitious interview:* This amounts to cheating by the interviewer. Here, the questionnaires are filled without conducting interviews. A surprise check by superiors is one way to minimise this.

(b) ***Office editing:*** Office editing is more thorough than field editing. The job of an office editor is more difficult than that of the field editor. In case of a mail questionnaire there are no other methods of cross-verification, except to conduct office audit. Examples as below illustrate the kind of problems faced by office editors. Problems of consistency, rapport with respondents, etc., are some of the issues which get highlighted during office editing.

Notes

Examples:

1. A respondent indicated that he doesn't drink coffee, but when questioned about his favourite brand, he replied "Bru".
2. A rating scale given to a respondent states Semantic Differential Scale with 10 items. The respondent has ticked "strongly agree" to the 10 items.
3. "What is the most expensive purchase you have made in the last one year?" is the question. Two respondents answer (1) LCD TV and (2) Trip to USA.

In example-1 above, there is inconsistency. There are two possibilities which an editor need to consider. (1) Was the respondent lying? (2) Did the interviewer record wrongly? The editor has to look into the answers to other questions on beverages, and interpret the right answer.

In example-2 above, it is to be remembered that Semantic Differential scale consists of items which has alternately positive and negative connotations. If a respondent has marked both positive and negative as 'agreed', the only conclusion the editor can draw is that the respondent filled the questionnaire without knowledge. The editor will have to discard this questionnaire, since there are no alternatives.

In example-3 above, both the respondents have answered correctly. The frame of reference is different. The main problem is, one of them is a product, whereas the other is a service. While coding the data, the two answers should be put under two different categories.

Answers to open-ended questions pose great difficulty in editing.

Coding

Coding refers to those activities which helps in transforming edited questionnaires into a form that is ready for analysis. Coding speeds up the tabulation while editing eliminates errors. Coding involves assigning numbers or other symbols to answers so that the responses can be grouped into limited number of classes or categories.

Example: 1 is used for male and 2 for female.

Some guidelines to be followed in coding which is as follows:

1. Establishment of appropriate category.
2. Mutual exclusivity.
3. Single Dimension.

1. ***Establishment of appropriate category:***

Example: Suppose the researcher is analysing the "inconvenience" that a car owner is facing with his present model. Therefore, the factor chosen for coding may be "inconvenience". Under this there could be 4 types (1) Inconvenience in entering the backseat (2) Inconvenience due to insufficient legroom (3) Inconvenience with respect to the interior (4) Inconvenience in door locking, and opening the dickey. Now the researcher may classify these four answers based on internal inconvenience and other inconveniences referring to the exterior. Each is assigned a different number for the purpose of codification.

2. ***Mutually exclusive:*** This is important because the answer given by the respondent should be placed under one category. Example: Occupation of an individual may be responded to as (1) Professional (2) Sales (3) Executive (4) Manager etc.

Notes

Sometimes, respondents might think that they belong to more than one category. This is because a sales personnel may be doing a sales job and therefore should be placed under the sales category. Also, he may supervise the work of other sales executive(s). In this case, he is doing a managerial function. Viewed in this context, he should be placed under the managerial category, which has a different code. Therefore, he can only be put under one category, which is to be decided. One way of deciding this could be to analyse "in which of the two functions does he spend most time"?

Yet another scenario assumes that there is a salesman who is currently employed. Under the column of 'occupation', he will tick it as sales, while under the current employment column, he will mark unemployed. How does one codify this? Under which category should he be placed. One of the solutions is to have a classification, such as employed salesman, unemployed salesman to represent the two separate categories.

Questions	Answers	Codes
1. Do you own a vehicle	Yes	1
	No	2
2. What is your occupation	Salaried	S
	Business	B
	Retired	R
	Technical	T
	Consultant	C

Tabulation

Tabulation refers to counting the number of cases that fall into various categories. The results are summarized in the form of statistical tables. The raw data is divided into groups and sub-group(s). The counting and placing of data in a particular group and sub-group are done. The tabulation involves:

(1) Sorting and counting

(2) Summarising of data

Tabulation may be of two types (1) simple tabulation (2) cross tabulation. In simple tabulation, a single variable is counted. Cross-tabulation includes two or more variables, which are treated simultaneously. Tabulation can be done entirely by hand or by machine, or by both hand and machine.

The form in which tabulation is to be done is decided by taking into account (1) the purpose of study and (2) the use of statistical tools e.g. mean, mode, standard deviation etc. Improper tabulation may create difficulties in the use of these tools.

Sorting and Counting of Data

Sorting by manual method is as follows:

Sorting of data

Income (₹)	Tally Mark	Frequencies
1,000	𝍸	5
1,500	𝍸 𝍸	8
2,000	𝍸 𝍸 𝍸	12
2,500	𝍸 𝍸 |||| |	16

Notes

The above method is used commonly for sorting of data.

The tabulation may include table number, title, head note, stub, caption, sub-entries, body of the table, footnote and the source. The following example explains the component of a table.

Format of a Blank Table

TABLE No.

Title – Number of children per family

Head Note – Unit of measurement

Sub Heading	Caption	Total
	Body	
	Foot note	

The table must have a clear and brief title. The head note, usually the measurement unit, is placed at the top of the table in the right hand corner in a bracket.

Stub indicates the row title or the row headings and is placed in the left-hand column. Caption indicates what each column is meant for.

Sub-entries are the sub-group of the stub. The body of the table given full information of the frequency.

Summarising the Data

Before taking up summarising, the data should be classified into (1) Relevant data, and (2) Irrelevant data. During the field study, the researcher collects lot of data which he may think would be of use. Summarizing the data includes:

Classification of Data

(a) ***Number of groups:*** The number of groups should be sufficient to record all possible data. The classification should not be too narrow. If it is too narrow, there can be an overlap.

Example: If a researcher is conducting a survey on "Why does the current owner dislikes the car?" The car owner may indicate the following:

(1) Difficulty in seeking entry to the back seat

(2) Interior space

(3) Cramped leg room

(4) Mileage

(5) Rattling of the engine

(6) Dickey space

Now the above data can be classified into two or three categories such as (1) Discomfort (2) Expense (3) Pride (4) Safety (5) Design of the car.

Notes

(b) ***Width of the class interval:*** Class interval should be uniform and should be of equal width. This will provide consistency in the data distribution.

(c) ***Exclusive categories:*** The classification should be done in such a way that the response can be placed in only one category.

Example: Problem of leg room is the answer by respondent. This should be placed either under discomfort or design, but not both.

(d) ***Exhaustive categories:*** This should be made to include all responses including "Don't Know" answers. Sometimes this will influence the ultimate answer to the research problem.

(e) ***Avoid extremes:*** Avoid open-ended class interval.

Usage of Statistical Tools

Frequency Distribution: Frequency distribution simply reports the number of responses that each question receives. Frequency distribution organises the data into classes or groups. It shows the number of data that falls into particular class.

Example:

Income	No. of people
4000-6999	100
7000-9999	122
10000-12999	140

In marketing research, central value or **tendency plays a very important role.** The researcher may be interested in the average sales/shop, average consumption per month etc. The population parameters can be calculated with the help of simple average. The average of sample may be taken as population parameter. For example, if the average income of the population is to be computed, the researcher may select a sample, collect data on family income and calculate the relevant statistics which will be a representative of the population.

The total purchasing power of the community can be estimated on sample average. If the sample is stratified, the purchasing power of each income class may also be estimated. The median figure will reveal that half the population has more income than the median income, and the others half has less income than the median income. The mode will reveal the most common frequency. Based on this, shoppers can devise their strategy to sell the product.

The three most common ways to measure centrality or **central tendency are the mode, median and mean.**

Mode

The mode is the central value or item that occurs most often, when data is categorized in a frequency distribution, it is very easy to identify the mode, since the category in which the mode lies has the greatest number of observations.

Example: Data regarding household income of 300 people as tabulated by the researcher.

Notes

Income (₹)	Number (f)	Cumulative Frequency
upto 10000	30	30
10000-14999	125	155
20000-24999	50	205
25000-29999	30	235
30000-34999	33	268
35000-49999	20	288
above 35000	12	300

In the above table, 125 is the modal class.

Mode can be calculated using the formula:

$$M_0 = LM_0\left[\frac{D_1}{D_1 + D_2}\right] \times i$$

LM_0 = Lower limit of modal class.

D_1 = Difference between the frequency of modal class and the class immediately preceding the modal class.

D_2 = Difference between the frequency of the modal class and the class immediately succeeding the modal class.

i = size of the modal class interval.

$$M_d = 10,000 + \left(\frac{95}{95+75}\right) \times 5,000$$

substitute the values

$$= 10000 + \left(\frac{95}{170}\right)$$

$$= 5000 = 1000+2794 = ₹\ 12794$$

Conclusion: The majority have the income of ₹ 12,794. This is how statistical techniques are used in MR application.

Median

Median lies precisely halfway between the highest and lowest values. It is necessary to arrange the data into ascending or descending order before selecting the median value. For the ungrouped data with an odd number of observations, the median would be the middle value. For an even number of observations, the median value is half way between central value.

For a grouped data, the median is calculated using the formula:

$$M_d = LM_d \frac{\left(\frac{N}{2} - C.F\right)}{fM_d} \times i$$

M_d = Lower limit of median class.

Notes

CF = Cumulative frequency for the class just below the median class.

FMd: Frequency of the median class.

i = Size of the class interval of median class.

In the tableN = 300 N/2 = 150. The class containing the 150th person is the median class.

Substitute the value, we get median Md = 21568.

Conclusion: Half of the population has income > ₹ 21,568' and half of the population has income < ₹ 21,568.

Mean

In a grouped data, the midpoint of each category would be multiplied by the number of observation in that category. Sum up and the total to be divided by the total number of observation.

Eqn., $$X = \sum\left(\frac{fx}{\sum f}\right)$$

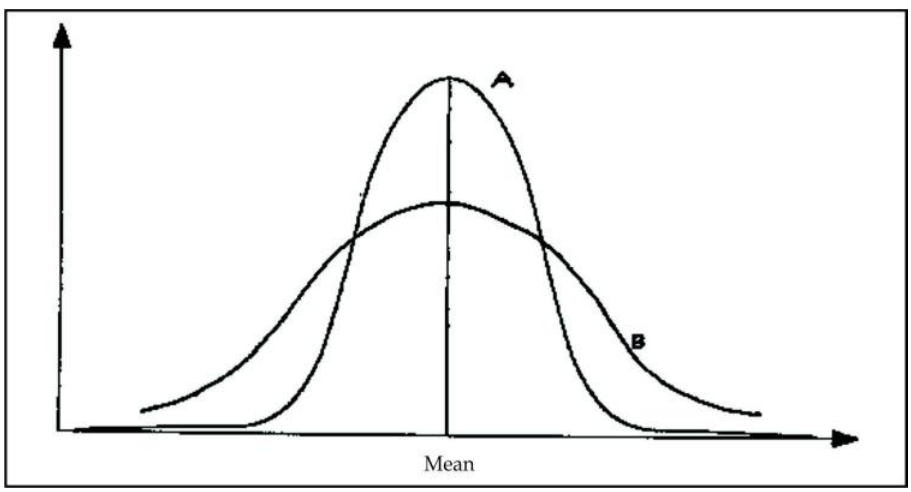
Example: Two students X, Y attend 3 class tests and the scores are as follows:

Marks		1st Test	2nd Test	3rd Test	Mean	
X		55%	60%	65%	60%	
Y		65%	60%	55%	60%	
Conclusion	X	- has improved				
	Y	- has deteriorated				

Though Mean is the same, X is better than Y.

Measures of Dispersion

Dispersion is the spread of the data in a distribution. A measure of dispersion:



Indicates the degrees of the scatteredness of the observations. Let curves A and B represent two frequency distributions. Observe that A and B have the same mean. But curve A has less variability than B.

Notes

If we measure only the mean of these two distributions, we will miss an important difference between A and B. To increase our understanding of the pattern of the data, we must also measure its dispersion.

Range: It is the difference between the highest and lowest observed values.

i.e range = H – L, H = Highest, L = Lowest.

Note:

1. Range is the crudest measure of dispersion.
2. $\frac{H-L}{H+L}$ is called the coefficient of range.

Semi-Inter Quartile Range (Quartile deviation): Semi-Inter quartile range Q.

Q is given by $$Q = \frac{Q_3 - Q_1}{2}$$

Note:

1. $\frac{Q_3 - Q_1}{Q_3 + Q_1}$ is called the **coefficient of quartile deviation.**
2. Quartile deviation is not a true measure of dispersion but only a distance of scale.

Mean Deviation (MD): If A is any average then mean deviation about A is given by:

$$MD(A) = \frac{\sum f_i \,|\, x_i - A \,|}{N}$$

Note:

1. Mean deviation about mean $MD(\bar{x}) = \frac{\sum f_i \,|\, x_i - \bar{x} \,|}{N}$
2. Of all the mean deviations taken about different averages mean derivation about the median is the least.
3. $\frac{MD(A)}{A}$ is called the coefficient of mean deviation.

Variance and Standard Deviation

Variance (σ^2): A measure of the average squared distance between the mean and each term in the population.

$$\sigma^2 = \frac{1}{N}\sum f_i (x_i - \bar{x})^2$$

Standard deviation (σ) is the positive square root of the variance:

$$\sigma = \sqrt{\frac{1}{N}\sum f_i (x_i - \bar{x})^2}$$

$$\sigma^2 = \frac{1}{N}\sum f_i (x_i^2 - (\bar{x})^2$$

Note: Combined variance of two sets of data of N_1 and N_2 items with means x_1 and x_2 and standard deviations σ^1 and σ^2 respectively is obtained by:

Notes

$$\sigma^2 = \frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_1 d_2^2}{N_1 + N_2}$$

Where $= (x - x_1)^2 \quad d_2^2 = (x - x_2)^2$

and $$x = \frac{N_1\overline{x_1} + N_2\overline{x_2}}{N_1 + N_2}$$

Sample variance **(σ^2)** *:* Let $x_1, x_2, x_3, \ldots\ldots x_n$, represent a sample with mean **x.**

Then sample variance σ^2 is given by:

$$\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

$$= \frac{\sum x^2}{n-1} - \frac{n(\bar{x})^2}{n-1}$$

Note: $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum x^2}{n-1} - \frac{n(\bar{x})^2}{n-1}}$ is called the sample standard deviation.

Task You have collected data on employees of a large organisation in a metro. You analyse the data by the type of work, education level, whether the employee belongs to an urban or rural area. The results are as below. How would you interpret them?

Annual Employee Turnover*

	Higher Education		Lower Education	
	Salaried monthly	**Daily wage**	**Salaried monthly**	**Daily wage**
Rural	6	14	18	18
Urban	10	12	19	20

*Turnover per 1,000 employees

Caselet

XYZ company is into pharmaceuticals to produce a medicine 'A', which is a pain reliever. A survey was conducted with doctors as sample and the following questions were asked.

"Would you recommend product 'A' to your patients when they suffer from pain"?

Yes _____ No _____

An analysis of the above showed that 75% of doctors surveyed said 'Yes', the rest said 'No'. From this survey, XYZ company made the following inference. "Three out of four doctors have recommended product 'A' for their patients, who suffer from pain".

Contd...

Notes

1. Is the inference valid?
2. If not, how else will you confirm that three out of four doctors have recommended this?

Self Assessment

Fill in the blanks:

13. ……of data is a method for deriving conclusions from the data analysed.
14. ……involves inspection and correction of each questionnaire.
15. …….editing is more thorough than field editing.
16. ……..refers to that activity which helps in transforming edited questionnaires into a form that is ready for analysis.
17. ……..refers to counting the number of cases that fall into various categories. The results are summarized in the form of statistical tables.

9.5 Summary

- Measurement can be made using nominal, ordinal, interval or ratio scale.
- These scales show the extent of likes/dislikes, agreement disagreement or belief towards an object.
- Each of the scale has certain statistical implications.
- There are four types of scales used in market research namely paired comparison, Likert, semantic differential and thurstone scale.
- Likert is a five point scale whereas semantic differential scale is a seven point scale.
- Bipolar adjectives are used in semantic differential scale.
- Thurstone scale is used to assess attitude of the respondents group regarding any issue of public interest.
- MDS uses perceptional map to evaluate customer's attitudes.
- The attribute or non-attribute method could be used.
- Validity and reliability of the scale is verified before the scale is used for measurement.
- If repeated measurement gives the same result, then the scale said to be reliable.
- Validity refers to "Does the scale measure what it intends to measure".
- There are three methods to check the validity which type of validity is required depends on "What is being measured".
- Processing data is very important in market research. After collecting the data, the next task of the researcher is to analyse and interpret the data.

9.6 Keywords

Constant Sum Scale: Constant sum scale is one of the methods of comparative scaling. In this method, the respondent is instructed to allocate some constant sum (points) to various features given, based on the importance of attribute to the respondent.

Notes

Interval Scale: Interval scale is more powerful than the nominal and ordinal scales. The distance given on the scale represents equal distance on the property being measured.

Likert scale: This consists of a series of statements concerning an attitude object. Each statement has '5 points', Agree and Disagree on the scale.

Multi-Dimensional Scaling: This is used to study consumer attitudes, particularly with respect to perceptions and preferences.

Nominal Scale: In this scale, numbers are used to identify the objects.

Ordinal Scale (Ranking scale): The Ordinal scale is used for ranking in most market research studies. Ordinal scales are used to ascertain the consumer perceptions, preferences, etc.

Rank Order Scale: In this method, respondents are required to rank more than two objects or alternatives based on some criteria.

Ratio Scale: Ratio scale is a special kind of internal scale that has a meaningful zero point. With this scale, length, weight or distance can be measured.

Scaling: The generation of a continuum upon which measured objects are located.

9.7 Review Questions

1. What are the four types of basic scales?
2. What is a paired comparison scale?
3. What is the statistical implication of various scales?
4. What are comparative and non comparative scales?
5. Explain the construction of :
   (a) Likert scale
   (b) Semantic differential scale
   (c) Thurstone scale
   (d) Juster scale
6. What is constant sum scale?
7. What is rank order scale?
8. Explain staple scale.
9. What is forced and unforced scale?
10. What is attribute and non attribute method in scaling?
11. What is M.D.S? And what are the limitations?
12. What are the different types, sources and characteristics of hypothesis?
13. A highway petrol police on NH4 wants to find out how fast the car and truck travels on this highway stretch. To obtain this information, a speed recording device at an appropriate place on the highway was installed. The speed was recorded for about three hours and the following data was recorded.

Notes

Speed in miles / hr.

73	49	70	63
55	61	60	68
52	50	69	60
65	66	59	62

Calculate the appropriate statistics for central tendency and dispersion.

14. The XYZ TV manufacturer is in the market for the past eight years. A survey conducted in the past by an MR agency produced the following score using *Likert Scale*. The data for various years is as below:

2000	-	18
2001	-	16
2002	-	17
2003	-	18
2004	-	20

(a) What do you conclude about the customers' attitude?

(b) Is it favourable or unfavourable?

15. A consumer durable company manufacturing TV sets is trying to measure consumer attitudes towards a product. For this purpose, the company wants the customer to complete a questionnaire which indicates several product attributes. It was decided by the company that only five attributes that affect the sale of the product would be considered for analysis. The attributes were appearance, quality of the picture, sound, after-sales service and price. The following scales are used to assess the product:

	Strongly disagree	**Disagree**	**Neither Agree nor Disagree**	**Agree**	**Strongly agree**
1. Appearance is good	___	___	___	___	X
2. Price is reasonable	___	___	___	X	___
3. After-sales service is good	___	___	___	X	___
4. Sound quality is excellent	___	X	___	___	___
5. Picture is sharp	___	___	___	___	X
	1	2	3	4	5

Suppose the customer has inspected the product and the response is as shown in the table above:

(a) What is the total score?

(b) What according to you is the attitude of the customer? Is it favourable or unfavourable?

Answers: Self Assessment

1. Interval
2. Thurstone
3. Multi-dimensional
4. Equal appearing interval

Notes

5. Likert
6. Item part, Evaluation part
7. Opinion
8. High
9. Accuracy, consistency
10. Reliability, validity
11. Semantic
12. Internal Consistency
13. Interpretation
14. Editing
15. Office
16. Coding
17. Tabulation

9.8 Further Readings

Books

Alan T Shao, *Marketing Research*, Cengage.

Cisnal Peter, *Marketing Research*, MCGE.

Hague & Morgan, *Marketing Research in Practice*, Kogan page.

Paneerselvam. R, *Research Methods*, PHI

Online links

www.indiastudychannel.com

www.scribd.com/doc

www.soas.ac.uk

www.web-source.net

Notes

Unit 10: Correlation

CONTENTS

Objectives
Introduction
10.1 Correlation
10.1.1 Scatter Diagram
10.2 Types of Correlation
10.2.1 Positive Correlation
10.2.2 Negative Correlation
10.2.3 No Correlation
10.3 Partial Correlation
10.4 Multiple Correlations
10.5 Summary
10.6 Keywords
10.7 Review Questions
10.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Learn about the concept of correlation;
- Identify the types of correlation;
- Explain the Karl Pearson's coefficient;
- Discuss the partial correlation;
- Describe the multiple correlation.

Introduction

Marketing research data analysis is a blend of statistics, psychology, information technology and art. The professional marketing researcher is not expected to have a complete understanding of all the techniques of data analysis, but is expected to manage the blending of these disciplines in order to develop and organize a complete analysis of the data that satisfies the information requirements of the project. Managers of today often need to understand and make decisions depending upon the numerical data on two or more variables simultaneously. For example,

(i) Cost of production and volume of production,

(ii) Expenditure on Advertising and Sales of a Product,

(iii) Number of Vehicles on Road and Number of Accidents,

(iv) Number of Colleges offering MBA Programme and number of MBA Graduates,

(v) Number of Counters at an e - Seva Kendra and the waiting time of customers

(vi) Number of Telephone calls and Rate per Call and so on.

Notes

In other words, one of the basic functions of a manager is to understand the relationship between these variables and make appropriate decisions keeping the future in mind known as 'Forecasting or Prediction'. The part which deals with understanding of the behaviour of variables is **Correlation** and the part deal with the forecasting is **Regression**.

10.1 Correlation

Correlation is a statistical tool for studying the relationship between two or more variables and correlation analysis involves various methods and techniques used for studying and measuring the extent of relationship between the two variables. Two variables said to be correlated, if the change in one variable results in a corresponding change in the other. Various experts have defined correlation in their own words and their definitions, broadly speaking, imply that correlation is the degree of association between two or more variables. Some important definitions of correlation are given below:

1. "If two or more quantities vary in sympathy so that movements in one tend to be accompanied by corresponding movements in other(s) then they are said to be correlated."

   — *L.R. Connor*

2. "Correlation is an analysis of covariation between two or more variables."

   — *A.M. Tuttle*

3. "When the relationship is of a quantitative nature, the appropriate statistical tool for discovering and measuring the relationship and expressing it in a brief formula is known as correlation."

   — *Croxton and Cowden*

4. "Correlation analysis attempts to determine the 'degree of relationship' between variables".

   — *Ya Lun Chou*

Correlation Coefficient: It is a numerical measure of the degree of association between two or more variables.

Did u know? **What is the scope of correlation analysis?**

The existence of correlation between two (or more) variables only implies that these variables (i) either tend to increase or decrease together or (ii) an increase (or decrease) in one is accompanied by the corresponding decrease (or increase) in the other. The questions of the type, whether changes in a variable are due to changes in the other, i.e., whether a cause and effect type relationship exists between them, are not answered by the study of correlation analysis. If there is a correlation between two variables, it may be due to any of the following situations:

1. ***One of the variable may be affecting the other:*** A correlation coefficient calculated from the data on quantity demanded and corresponding price of tea would only reveal that the degree of association between them is very high. It will not give us any idea about whether price is affecting demand of tea or vice-versa. In order to know this, we need to have some additional information apart from the study of correlation. For example if, on the basis of some additional information, we say that the price of tea affects its demand, then price will be the cause and quantity will be the effect. The causal variable is also termed as independent variable while the other variable is termed as dependent variable.

Notes

2. ***The two variables may act upon each other:*** Cause and effect relation exists in this case also but it may be very difficult to find out which of the two variables is independent. For example, if we have data on price of wheat and its cost of production, the correlation between them may be very high because higher price of wheat may attract farmers to produce more wheat and more production of wheat may mean higher cost of production, assuming that it is an increasing cost industry. Further, the higher cost of production may in turn raise the price of wheat. For the purpose of determining a relationship between the two variables in such situations, we can take any one of them as independent variable.

3. ***The two variables may be acted upon by the outside influences:*** In this case we might get a high value of correlation between the two variables, however, apparently no cause and effect type relation seems to exist between them. For example, the demands of the two commodities, say X and Y, may be positively correlated because the incomes of the consumers are rising. Coefficient of correlation obtained in such a situation is called a spurious or nonsense correlation.

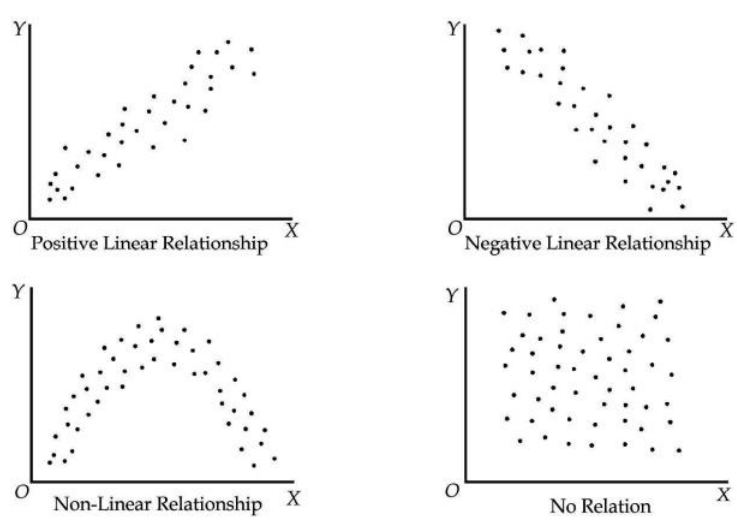
4. ***A high value of the correlation coefficient may be obtained due to sheer coincidence (or pure chance):*** This is another situation of spurious correlation. Given the data on any two variables, one may obtain a high value of correlation coefficient when in fact they do not have any relationship. For example, a high value of correlation coefficient may be obtained between the size of shoe and the income of persons of a locality.

10.1.1 Scatter Diagram

Let the bivariate data be denoted by (X_i, Y_i), where $i = 1, 2 \ldots\ldots n$. In order to have some idea about the extent of association between variables X and Y, each pair (X_i, Y_i), $i = 1, 2\ldots\ldots n$, is plotted on a graph. The diagram, thus obtained, is called a Scatter Diagram.

Each pair of values (X_i, Y_i) is denoted by a point on the graph. The set of such points (also known as dots of the diagram) may cluster around a straight line or a curve or may not show any tendency of association. Various possible situations are shown with the help of following diagrams:

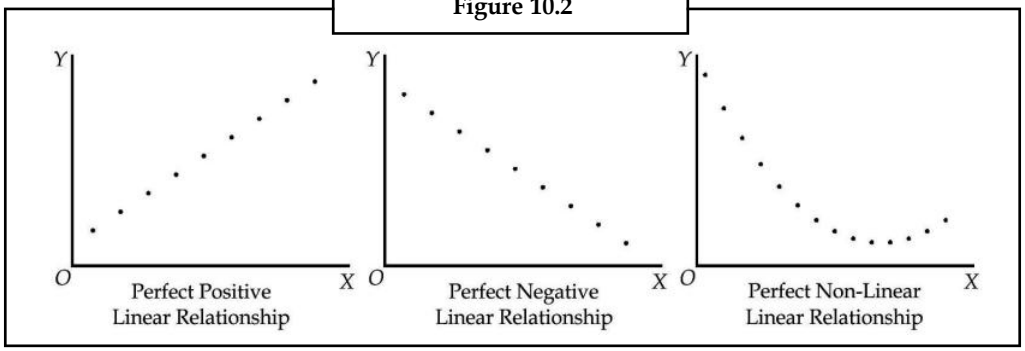
Figure 10.1: Scatter Diagram



Notes

If all the points or dots lie exactly on a straight line or a curve, the association between the variables is said to be perfect. This is shown below:

Figure 10.2



A scatter diagram of the data helps in having a visual idea about the nature of association between two variables. If the points cluster along a straight line, the association between variables is linear. Further, if the points cluster along a curve, the corresponding association is non-linear or curvilinear. Finally, if the points neither cluster along a straight line nor along a curve, there is absence of any association between the variables.

It is also obvious from the above figure that when low (high) values of X are associated with low (high) value of Y, the association between them is said to be positive. Contrary to this, when low (high) values of X are associated with high (low) values of Y, the association between them is said to be negative.

10.2 Types of Correlation

Broadly speaking, there are four types of Correlation, namely, (a) Positive correlation, (b) Negative correlation, (c) Linear correlation and (d) Non-Linear Correlation.

10.2.1 Positive Correlation

If the values of two variables deviate in the same direction i.e., if increase in the values of one variable results, on an average, in a corresponding increase in the values of the other variable or if a decrease in the values of one variable results, on an average, in a corresponding decrease in the values of the other variable, the corresponding correlation is said to be positive or direct.

Examples:

(i) Sales revenue of a product and expenditure on Advertising.

(ii) Amount of rain fall and yield of a crop (up to a point).

(iii) Price of a commodity and quantity of supply of a commodity.

(iv) Height of the Parent and the height of the Child.

(v) Number of patients admitted into a Hospital and Revenue of the Hospital.

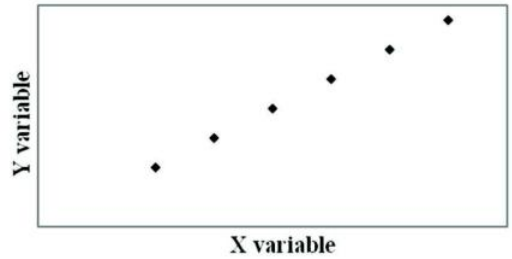
(vi) Number of workers and output of a factory.

Perfect Positive Correlation

If the variables X and Y are perfectly positively related to each other then, we get a graph as shown in Figure 10.3.

Notes

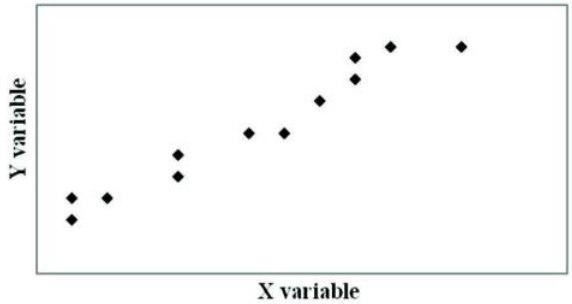
Figure 10.3: Perfect Positive Correlation (R = 1)



Very High Positive Correlation

If the variables X and Y are related to each other with a very high degree of positive relationship then we can notice a graph as in Figure 10.4.

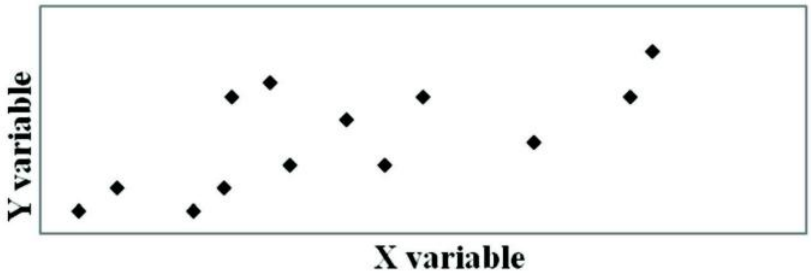
Figure 10.4: Very High Positive Correlation (R = Nearly +1)



Very Low Positive Correlation

If the variables X and Y are related to each other with a very low degree of positive relationship then we can notice a graph as in Figure 10.5.

Figure 10.5: Very Low Positive Correlation (R = near to +0)



10.2.2 Negative Correlation

Correlation is said to be negative or inverse if the variables deviate in the opposite direction i.e., if the increase (decrease) in the values of one variable results, on the average, in a corresponding decrease (increase) in the values of the other variable.

Notes

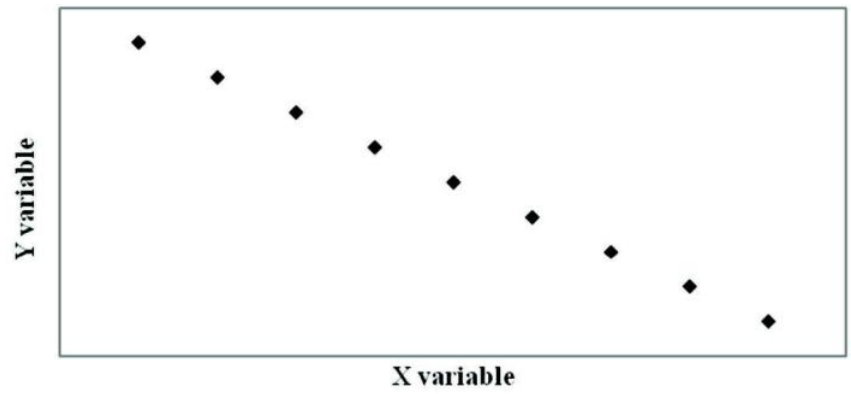
Examples:

1. Price and demand of a commodity
2. Sale of Woolen garments and the day temperature.

Perfect Negative Correlation

If the variables X and Y are perfectly negatively related to each other then, we get a graph as shown in Figure 10.6.

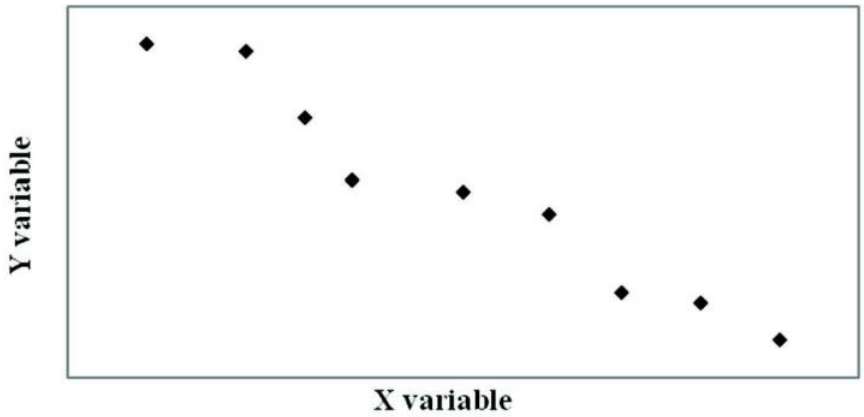
Figure 10.6: Perfect Negative Correlation (R = - 1)



Very High Negative Correlation

If the variables X and Y are related to each other with a very high degree of negative relationship then we can notice a graph as in Figure 10.7.

Figure 10.7 : Very High Negative Correlation (R = Near To - 1)

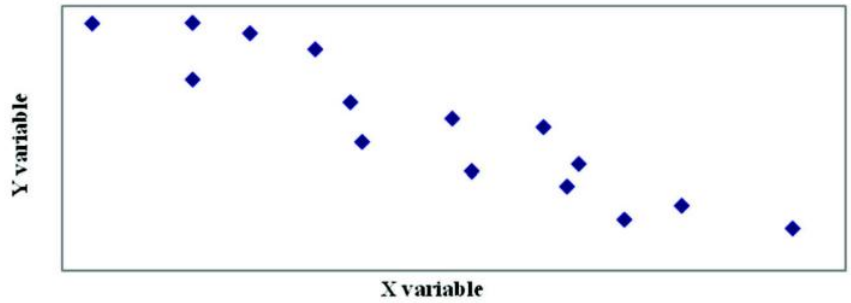


Very Low Negative Correlation

If the variables X and Y are related to each other with a very low degree of negative relationship then we can notice a graph as in Figure 10.8.

Notes

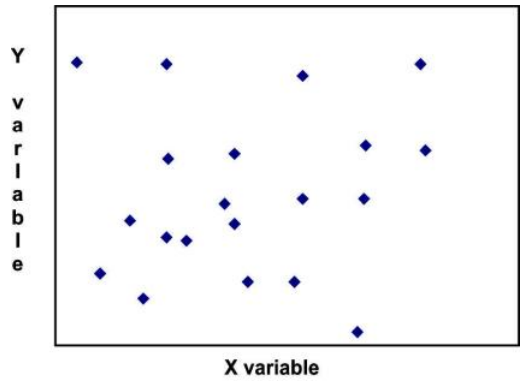
Figure 10.8 : Very Low Negative Correlation (R = Near to 0 but Negative)



10.2.3 No Correlation

If the scatter diagram show the points which are highly spread over and show no trend or patterns we can say that there is no correlation between the variables. Refer to Figure 10.9.

Figure 10.9: No Correlation (r = 0)



Linear Correlation

Two variables are said to be linearly related if corresponding to a unit change in one variable there is a constant change in the other variable over the entire range of the values.

If two variables are related linearly, then we can express the relationship as

$$Y = a + b X$$

where **'a'** is called as the **"intercept"** (If X= 0, then Y= a) and **'b'** is called as the **"rate of change"** or **slope.**

If we plot the values of X and the corresponding values of Y on a graph, then the graph would be a straight line as shown in Figure 10.10.

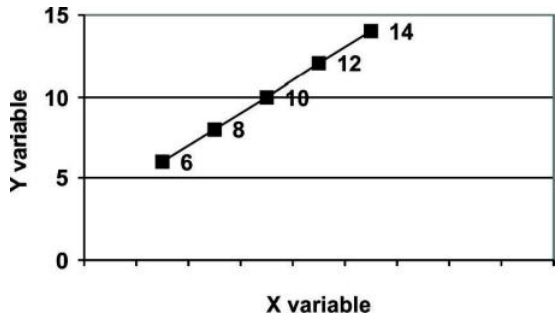
Example:

X	1	2	3	4	5
Y	6	8	10	12	14

Notes

For a unit change in the value of x, a constant 2 units change in the value of y can be noticed. The above can be expressed as: Y = 4 + 2 X

Figure 10.10



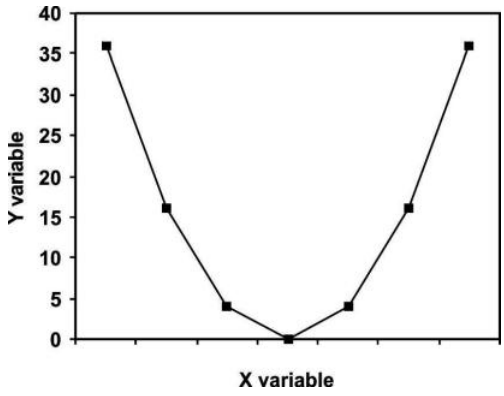
Non Linear (Curvilinear) Correlation

If corresponding to a unit change in one variable, the other variable does not change in a constant rate, but change at varying rates, then the relationship between two variables is said to be non-linear or curvilinear as shown in Figure 10.11. In this case, if the data are plotted on the graph, we do not get a straight line curve. Mathematically, the correlation is non-linear if the slope of the plotted curve is not constant. Data relating to Economics, Social Science and Business Management do exhibit often non-linear relationship. We confine ourselves to linear correlation only.

Example:

X	-6	-4	-2	0	2	4	6
Y	36	16	4	0	4	16	36

Figure 10.11: Non-linear Correlation



Karl Pearson's Coefficient of Correlation

To measure the degree of association between two variables X and Y, Karl Pearson defined the Coefficient of Correlation 'γ' as below. In this method, the coefficient of correlation is calculated as the ratio of the covariance of the two variables to the product of their variances.

Notes

$$\text{Correlation co-efficient } (\gamma) = \frac{Cov(X_i, Y_i)}{\sqrt{\{V(X_i)V(Y_i)\}}}$$

$$\text{where } Cov(X_i, Y_i) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

$$\text{and } V(X_i) = \frac{\sum(X_i - \bar{X})^2}{n}$$

$$V(Y_i) = \frac{\sum(Y_i - \bar{Y})^2}{n}$$

$$\gamma = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}}$$

Where $x = X_i - \bar{X}$ and $y = Y_i - \bar{Y}$

Pearson's Method – Direct Method

We apply direct formula to find Karl Pearson's Co-efficient of correlation.

Example 1: Following is the data on two variables X_i and Y_i we find the sums and squares of products as shown in the table below:

X_i	Y_i	$(X_i - \bar{X})$ x	$(Y_i - \bar{Y})$ y	$(X_i - \bar{X})^2$ x^2	$(Y_i - \bar{Y})^2$ y^2	$(X_i - \bar{X})(Y_i - \bar{Y})$ xy
2	7	-2	- 4	4	16	8
3	9	-1	- 2	1	4	2
4	10	0	- 1	0	1	0
5	14	1	3	1	9	3
6	15	2	4	4	16	8
20	55			$\sum x^2 = 10$	$\sum y^2 = 46$	$\sum xy = 21$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{20}{5} = 4, \bar{Y} = \frac{\sum Y_i}{n} = \frac{55}{5} = 11$$

$$\gamma = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}}$$

$$\gamma = \frac{21}{\sqrt{10}\sqrt{46}} = \frac{21}{3.16 \times 6.78} = 0.98$$

The value of $\gamma = 0.98$ shows that two series X and Y have almost perfect positive correlation.

Notes

Karl Pearson's Method – Without Deviations (Short-cut Method)

When the arithmetic means of both sets of numerical items are not whole numbers and involve decimals, calculating the coefficient of correlation by direct method becomes tedious. To overcome this difficulty the following modified short-cut method formula is used:

$$\text{Cov}(X_i, Y_i) = \frac{\sum X_i Y_i}{n} - \bar{X}\,\bar{Y}$$

$$V(X_i) = \frac{\sum X_i^2}{n} - \bar{X}^2 \; ; \; V(Y_i) = \frac{\sum Y_i^2}{n} - \bar{Y}^2$$

$$\gamma = \frac{\text{Cov}(X_i, Y_i)}{\sqrt{\{V(X_i)V(Y_i)\}}}$$

$$\gamma = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{\left[n\sum X_i^2 - \left(\sum X_i\right)^2\right]\left[n\sum Y_i^2 - \left(\sum Y_i\right)^2\right]}}$$

Example 2: Calculate the Karl Pearson's coefficient of correlation for the following data between sales and advertising expenditure.

Let sales represents Xi variable and advertise expenditure represents Yi variable to calculate the correlation coefficient using the following formula:

$$\gamma = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{\left[n\sum X_i^2 - \left(\sum X_i\right)^2\right]\left[n\sum Y_i^2 - \left(\sum Y_i\right)^2\right]}}$$

X_i	Y_i	X_i^2	Y_i^2	X_iY_i
1	3	1	9	3
2	15	4	225	30
3	6	9	36	18
4	20	16	400	80
5	9	25	81	45
6	25	36	625	150
$\sum X_i = 21$	$\sum Y_i = 78$	$\sum X_i^2 = 91$	$\sum Y_i^2 = 1376$	$\sum X_iY_i = 326$

$$\gamma = \frac{(6 \times 326) - (21 \times 78)}{\sqrt{\left[(6 \times 91) - (21)^2\right]\left[(6 \times 1376) - (78)^2\right]}}$$

$$\gamma = \frac{318}{(10.247 \times 46.605)}$$

$$\gamma = 0.667$$

This suggests that a fairly high degree of correlation between X and Y series i.e. between sales and advertising expenditure.

Notes

Karl Pearson's Method – Shifting Origin

In case the magnitude of the data is large, using the two methods explained above will give lot of inconvenience while calculating the correlation coefficient by Karl Pearson's method. So we take deviations from some convenient numbers to reduce the magnitude of data. There will be no change in the value of correlation coefficient even if deviations are taken. We define, $u_i = X_i$- A and = $v_i = Y_i$- B, where A and B can any arbitrary and assumed values. The formulae are given below,

$$V(u_i) = \frac{\sum u_i^2}{n} - \bar{u}^2 \text{ ; } V(vi) = \frac{\sum v_i^2}{n} - \bar{v}^2 \text{ ; } Cov(u_i,v_i) = \frac{\sum u_i v_i}{n} - \bar{u}\,\bar{v}$$

$$\gamma = \frac{Cov(u_i, v_i)}{\sqrt{\{V(u_i)\ V(v_i)\}}}$$

$$\gamma = \frac{n\sum u_i v_i - \sum u_i \sum v_i}{\sqrt{\left[n\sum u_i^2 - \left(\sum u_i\right)^2\right]\left[n\sum v_i^2 - \left(\sum v_i\right)^2\right]}}$$

Example 3: Using short cut method, we calculate 'r' for the following data of X_i = Advertising expenditure (Rupees in thousands) and Y_i = sales (Rupees in lakhs). Let us define A = 60 and B=70, two variables chosen arbitrarily. Then $u_i = X_i$ - 60 and $v_i = Y_i$ - 70

X_i	Y_i	u_i	v_i	u_i^2	v_i^2	$u_i v_i$
39	47	- 21	-23	441	529	+483
65	53	5	-17	25	289	- 85
62	58	2	-12	4	144	- 24
90	86	30	16	900	256	+480
82	62	22	- 8	484	64	-176
75	68	15	- 2	225	4	- 30
25	60	-35	-10	1225	100	-350
98	91	38	21	1444	441	+798
36	51	-24	-19	576	361	+456
78	84	18	14	324	196	+252
Total		$\sum u_i$ =50	$\sum v_i$ = -40	$\sum u_i^2$ = 5648	$\sum v_i^2$ = -2384	$\sum u_i v_i$ = 2504

$$\bar{u} = \frac{\sum u_i}{n} = \frac{50}{10} = 5; \bar{v} = \frac{\sum v_i}{n} = \frac{-40}{10} = -4$$

$$\gamma = \frac{n\sum u_i v_i - \sum u_i \sum v_i}{\sqrt{\left[n\sum u_i^2 - \left(\sum u_i\right)^2\right]\left[n\sum v_i^2 - \left(\sum v_i\right)^2\right]}}$$

$$\gamma = \frac{(10 \text{ x } 2540) - (50 \text{ x} - 40)}{\sqrt{\left[(10 \text{ x } 5648) - (50)^2\right]\left[(10 \text{ x } 2384) - (-40)^2\right]}}$$

$$\gamma = \frac{27040}{\sqrt{(53980)(22240)}} = \frac{27040}{34647.373}$$

$$\gamma = 0.78$$

Hence the correlation between X and Y series is fairly high as the coefficient of correlation is 0.78.

Notes

Properties of Correlation:

(i) The value of correlation coefficient γ varies between [–1, +1]. This indicates that the value of does not exceed unity.

(ii) Sign of γ depends on sign of the covariance.

(iii) If γ = –1, the variables are perfectly negatively correlated.

(iv) If γ = +1, the variables are perfectly positively correlated.

(v) If γ = 0, the variables are not correlated in a linear fashion. There may be nonlinear relationship between variables.

(vi) Correlation coefficient is independent of change of scale and shifting of origin. In other words, shifting the origin and change the scale do not have any effect on the value of correlation.

Let us see the following example to understand the concept, 'if γ = 0, the variables are not correlated in a linear fashion. There may be nonlinear relationship between variables'.

Example 4: If X_i and Y_i are given as below, we calculate the correlation coefficient.

X_i	Y_i	X_i^2	Y_i^2	X_iY_i
-3	9	9	81	-27
-2	4	4	16	-8
-1	1	1	1	-1
0	0	0	0	0
1	1	1	1	1
2	4	4	16	8
3	9	9	81	27
$\sum X_i = 0$	$\sum Y_i = 28$	$\sum X_i^2 = 28$	$\sum Y_i^2 = 196$	$\sum X_i^2 Y_i^2 = 0$

$$\gamma = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{\sqrt{\left[n\sum X_i^2 - (\sum X_i)^2\right]\left[n\sum Y_i^2 - (\sum Y_i)^2\right]}}$$

$$\gamma = \frac{(7 \times 0) - (0 \times 28)}{\sqrt{\left[(7 \times 28) - (0)^2\right]\left[(7 \times 196) - (28)^2\right]}}$$

$$\gamma = \frac{0}{\sqrt{196 \times 588}}$$

$$\gamma = \frac{0}{\sqrt{196 \times 588}} = 0$$

Since γ = 0 it does not mean that the variables X_i and Y_i are uncorrelated. It can only be said that the variables are linearly uncorrelated. In fact if we closely look at the data of X_i and Y_i, it can be observed that $Y_i = X_i^2$ is the relationship existing between X_i and Y_i. This is a nonlinear relationship between the variables. Karl Pearson's coefficient of correlation can not measure nonlinear relationship between the variables.

Notes

Correlation of Grouped Data

When the number of observations is large, the data are often classified into two-way frequency distribution i.e. table where in the values of one variable (X) are represented in the rows while other variable (Y) in columns. These values can be either discrete or continuous. The frequencies in each class are shown in cells in the body of the table.

Steps for calculating correlation coefficient for grouped data:

(i) Record the mid-points (mp) of the class intervals for both X and Y variables.

(ii) Choose an assumed mean in X series and calculate the deviations (dx) from it. The same procedure to be used for Y series and calculate the deviations (dy).

(iii) To simplify the calculations, step deviations can be taken by dividing deviations by a common factor.

(iv) Calculate f.dx, f .dx.dx i.e.f.dx^2, f.dx.dy for X series and f.dy, f .dy.dy i.e.f.dy^2, f.dx.dy for Y series.

(v) Substitute all the values obtained in the following formula:

$$\gamma = \frac{n\sum fdxdy - \sum fdx \sum fdy}{\sqrt{\left[n\sum fdx^2 - \left(\sum fdx\right)^2\right]\left[n\sum fdy^2 - \left(\sum fdy\right)^2\right]}}$$

Example 5: Calculate the Karl Pearson's coefficient of correlation for the following grouped data:

Sales Revenue (₹ lakh)	Advertising Expenditure (₹ lakh) 5 -15	15 - 25	25 - 35	35 -45	**Total**
75 - 125	3	4	4	8	**19**
125 - 175	8	6	5	7	**26**
175 - 225	2	2	3	4	**11**
225 - 275	3	3	2	2	**10**
Total	**16**	**15**	**14**	**21**	**66**

Y \ X			5-15	15-25	25-35	35-45				
		mp	10	20	30	40				
	mp	dy \ dx	-1	0	1	2	**f**	fdy	fdy^2	fdxdy
75-125	100	-1	3(3)	4(0)	4(-4)	8(-16)	**19**	-19	19	-17
125-175	150	0	8(0)	6(0)	5(0)	7(0)	**0**	0	0	0
175-225	200	1	2(-2)	2(0)	3(3)	4(8)	**11**	11	11	9
225-275	250	2	3(-6)	3(0)	2(4)	2(8)	**10**	20	40	6
Total		**f**	**16**	**15**	**14**	**21**	**66**	$\sum$ fdy **=12**	$\sum$ fdy^2 **= 70**	$\sum$ fdxdy **= -2**

Notes

fdx	-16	0	14	$\sum$ f.dx = 42
f.dx²	16	0	14	$\sum$ fdx² = 84
f.dxdy	-5	0	3	$\sum$ fdxdy = 0

The value in the bracket in each cell shows fdxdy

$$\gamma = \frac{n\sum fdxdy - \sum fdx \sum fdy}{\sqrt{\left[n\sum fdx^2 - \left(\sum fdx\right)^2\right]\left[n\sum fdy^2 - \left(\sum fdy\right)^2\right]}}$$

$$\gamma = \frac{(66 \times -2) - (40 \times 12)}{\sqrt{\left[66 \times 114 - (40)^2\right]\left[66 \times 70 - (12)^2\right]}}$$

$$\gamma = \frac{-9.27}{\sqrt{[89.76]\ [67.82]}}$$

$$\gamma = \frac{-9.27}{9.47 \times 8.24} = -0.119$$

This shows very low degree of negative correlation between advertising expenditure (X) and sales revenue (Y)

Rank Correlation (Spearman's Method)

It is not possible to express attributes such as character, conduct, honesty, beauty, morality, intellectual integrity etc. in numerical terms. For example, it is easy to for a class teacher to arrange the students in his class in an ascending or descending order of intelligence. This means that he can rank them according to their intelligence. Hence in problems that involve attributes of the type mentioned above, the coefficient of correlation is entirely based on the rank differences between corresponding items.

We may have two types of numerical problems in rank correlation:

(a) When actual ranks are given

(b) When ranks are not given

Calculation of Rank Correlation

(i) In the first case, when actual ranks are given, the difference of the two ranks ($R_1 - R_2$) are taken and these are denoted by 'd'

(ii) The differences are squared and their total (Σd^2) obtained

(iii) Then the following formula is applied to calculate the rank correlation coefficient

$$r_s = 1 - \frac{6\sum d^2}{N(N^2 - 1)}$$

Where r_s denotes Spearman's Rank Correlation and N denotes number of pairs of observations.

Notes

(iv) In the second case,1 when the ranks are not given, when the actual data are given, we have to assign ranks. We may do so by taking highest value as 1 or the lowest value as 1. When the two observations are same, then the normal practice is to assign an average rank to the two observations.

When the ranks are given:

Example 6: The ranking of 10 students in two subjects A and B are as follows:

Student	1	2	3	4	5	6	7	8	9	10
Ranks in Subject A	4	6	1	3	9	7	10	2	8	5
Ranks in Subject B	5	8	3	1	7	6	9	2	10	4

Calculate coefficient of rank correlation and comment on the result

Solution:

In order to calculate rank correlation, we have to calculate Σd^2 and the following formula is used

$$r_s = 1 - \frac{6\sum d^2}{N(N^2-1)}$$

The following table shows the calculations:

Student No.	**Ranks in Subject A (R_1)**	**Ranks in Subject B (R_2)**	**Difference ($R_1 - R_2$) (d)**	**Squared difference (d^2)**
1	4	5	-1	1
2	6	8	-2	4
3	1	3	-2	4
4	3	1	2	4
5	9	7	2	4
6	7	6	1	1
7	10	9	1	1
8	2	2	0	0
9	8	10	-2	4
10	5	4	1	1
				$\Sigma d^2 = 24$

$$r_s = 1 - \frac{6\sum d^2}{N(N^2-1)}$$

$$r_s = 1 - \frac{(6 \times 24)}{(10)(10^2-1)}$$

$$r_s = 1 - \frac{144}{(10)(99)} = 0.855$$

The rank correlation coefficient (0.855) shows that there is a very high degree of correlation between ranks obtained in subject A and Subject B of the ten students.

Notes

When the ranks are not given:

Example 7: Compute the Spearman's coefficient of correlation between marks assigned to ten students by Judges X and Y in a certain competitive test as shown below:

Student No	1	2	3	4	5	6	7	8	9	10
Marks by Judge X	43	56	29	81	96	34	73	62	48	76
Marks by Judge Y	15	26	34	86	19	29	83	67	51	58

Student No	**Marks by Judge X**	**Ranks by Judge X (R_1)**	**Marks by Judge Y**	**Ranks by Judge Y (R_2)**	**Difference ($R_1 - R_2$) (d)**	**Squared difference (d^2)**
1	43	8	15	10	-2	4
2	56	6	26	8	-2	4
3	29	10	34	6	4	16
4	81	2	86	1	1	1
5	96	1	19	9	-8	64
6	34	9	29	7	2	4
7	73	4	83	2	2	4
8	62	5	67	3	2	4
9	48	7	51	5	2	4
10	76	3	58	4	-1	1
						$\Sigma d^2 = 106$

$$r_s = 1 - \frac{6\sum d^2}{N(N^2-1)}$$

$$r_s = 1 - \frac{(6 \times 106)}{(10)(10^2-1)}$$

$$r_s = 1 - \frac{636}{(10)(99)} = 0.36$$

The rank correlation coefficient (0.36) shows that there is a low degree of correlation between marks assigned by Judge X and Judge Y to the ten students.

Example 8: Obtain the rank correlation between variables Xth (Price of commodity A in ₹) and Yth (Price of commodity B in ₹) from the following pairs of observed values.

X	24	29	23	38	46	52	41	36	68	56
Y	110	126	145	131	163	158	131	129	154	140

X	**Ranks of X (R_1)**	**Y**	**Ranks of Y (R_2)**	**Difference ($R_1 - R_2$) (d)**	**Squared difference (d^2)**
24	9	110	10	-1	1
29	8	126	9	-1	1
23	10	145	4	6	36
38	6	131	6.5	-0.5	0.25

Contd...

Notes

52	3	158	2	1	1
41	5	131	6.5	-1.5	2.25
36	7	129	8	-1	1
68	1	154	3	-2	4
56	2	140	5	-3	9
					$\Sigma d^2 = 64.5$

In the data, there two equal values (found in Y series) i.e. 131 which is a tie for the ranks 6 and 7 respectively. Then the average of 6 and 7 ranks (6.5) is assigned as rank for both the observations. Then the common ranks for both the observations are 6.5.

In this data we find common ranks in the second series (Y). Therefore the formula for the coefficient of correlation through the rank differences method has to be modified as given below:

$$r_s = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3) + ...\right]}{N(N^2 - 1)}$$

m_1, m_2, m_3 …. stands for number of items in the respective groups with common ranks. In this problem only one group having items two (or two common ranks in that group), hence we can assign $m_1 = 2$

$$r_s = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(m_1^3 - m_1)\right]}{N(N^2 - 1)}$$

$$r_s = 1 - \frac{6\left[64.5 + \frac{1}{12}(2^3 - 2)\right]}{10(10^2 - 1)}$$

$$r_s = 1 - \frac{6[64.5 + 0.5]}{990} = 0.61$$

The rank correlation coefficient (0.61) shows that there is a moderate correlation between X and Y.

Self Assessment

Fill in the blanks:

1. The coefficient of correlation obtained on the basis of ranks is called .....
2. The only merit of Karl Pearson's coefficient of correlation is that it is the most popular method for expressing the ..... and ..... of linear association.
3. The ..... of correlation coefficient is an amount which if added to and subtracted from the mean correlation coefficient, gives limits within which the chances are even that a coefficient of correlation from a series selected at random will fall.
4. The value of Karl Pearson's coefficient is unduly affected by ..... items.

10.3 Partial Correlation

In case of three variables x_i, x_j and x_k, the partial correlation between x_i and x_j is defined as the simple correlation between them after eliminating the effect of xk. This is denoted as $r_{ij} \times k$.

Notes

We note that $x_i \times k = x_i - bikxk$ is that part of xi which is left after the removal of linear effect of x_k on it. Similarly, $x_j \times k = x_j - bjkxk$ is that part of xj which is left after the removal of linear effect of xk on it. Equivalently, $r_{ij} \times k$ can also be regarded as correlation between $x_i \times k$ and $x_j \times k$. Thus, we can write .

Using property III of residual products, we can write $r_{ij \cdot k} = \dfrac{\sum x_{i \cdot k} x_{j \cdot k}}{\sqrt{\sum x_{i \cdot k}^2 x_{j \cdot k}^2}}$

$$\begin{aligned} Sx_{i\times k}x_{j\times k} &= S_{xi\times kx}j = S(x_i - b_{ik}x_k)x_j = Sx_ix_j - b_{ik}Sx_jx_k \\ &= nS_iS_jr_{ij} - r_{ik}\frac{S_i}{S_k}nS_jS_kr_{jk} = nS_iS_j\left(r_{ij} - r_{ik}r_{jk}\right) \end{aligned}$$

Further, using property III, we can write

$$\begin{aligned} \Sigma x_{i\times k}^2 &= Sxixi\times k = Sxi(xi - bikxk) = Sxi^2 - bikSxixk \\ &= nS_i^2 - r_{ik}\frac{S_i}{S_k}nS_iS_kr_{ik} = nS_i^2\left(1 - r_{ik}^2\right) \end{aligned}$$

Similarly, $\Sigma x_{i\times k}^2 = nS_j^2\left(1 - r_{jk}^2\right).$

Thus, we have $$ri_{j\times k} = \frac{nS_iS_j\left(r_{ij} - r_{ik}r_{jk}\right)}{\sqrt{nS_i^2\left(1 - r_{ik}^2\right)nS_j^2\left(1 - r_{jk}^2\right)}} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{\left(1 - r_{ik}^2\right)\left(1 - r_{jk}^2\right)}}$$

Did u know? **What is Zero order, First order, and Second order Partial Correlation?**

Simple correlation between two variables is called the zero order co-efficient since in simple correlation, no factor is held constant. The partial correlation studied between two variables by keeping the third variable constant is called a first order co-efficient, as one variable is kept constant. Similarly, we can define a second order co-efficient and so on. The partial correlation co-efficient varies between –1 and +1. Its calculation is based on the simple correlation co-efficient.

10.4 Multiple Correlations

The coefficient of multiple correlations in case of regression of xi on xj and xk, denoted by $Ri\times jk$, is defined as a simple coefficient of correlation between xi and xic.

Thus $$R_{i\times jk} = \frac{Cov\left(x_i, x_{ic}\right)}{\sqrt{Var\left(x_i\right)Var\left(x_{ic}\right)}} = \frac{\sum x_ix_{ic}}{\sqrt{\sum x_i^2 \sum x_{ic}^2}} = \frac{\sum x_i\left(x_i - x_{i.jk}\right)}{\sqrt{\sum x_i^2 \sum\left(x_i - x_{i.jk}\right)^2}}$$

$$= \frac{\sum x_i^2 - \sum x_ix_{i\cdot jk}}{\sqrt{\sum x_i^2 \sum\left(x_i - x_{i\cdot jk}\right)x_i}} = \frac{\sum x_i^2 - \sum x_ix_{i\cdot jk}}{\sqrt{\sum x_i^2\left(\sum x_i^2 - \sum x_ix_{i\cdot jk}\right)}} \quad \text{(Using property III)}$$

$$= \frac{nS_i^2 - nS_{i.jk}^2}{\sqrt{nS_i^2\left(nS_i^2 - nS_{i.jk}^2\right)}} = \frac{1}{S_i}\sqrt{S_i^2 - S_{i.jk}^2}$$

Notes

Square of $R_{i\times jk}$ is known as the coefficient of multiple determination.

$$R^2_{i\cdot jk} = \frac{1}{S_i^2}\left(S_i^2 - S^2_{i\cdot jk}\right) = 1 - \frac{S^2_{i\cdot jk}}{S_i^2}$$

It may be noted here that $\frac{S^2_{i\cdot jk}}{S_i^2}$ is proportion of unexplained variation. Thus, we can also write

$R^2_{i\cdot jk} = 1 - \frac{x^2_{i\cdot jk}}{x_i^2}$.

Further, we can write $R^2_{i\cdot jk}$ in terms of the simple correlation coefficients.

$$R^2_{i\cdot jk} = 1 - \frac{S_i^2\left(1 - r_{ij}^2 - r_{ik}^2 - r_{jk}^2 + 2r_{ij}r_{ik}r_{jk}\right)}{S_i^2\left(1 - r_{jk}^2\right)} = \frac{r_{ij}^2 + r_{ik}^2 - 2r_{ij}r_{ik}r_{jk}}{1 - r_{jk}^2}$$

Notes If there are m variables, $R^2_{1\cdot 23....m} = 1 - \frac{S^2_{1\cdot 23....m}}{S_1^2} = 1 - \frac{\sum x^2_{1\cdot 23....m}}{\sum x_1^2}$

Coefficient of Multiple Correlations

The multiple correlation coefficient generalizes the standard coefficient of correlation. It is used in multiple regression analysis to assess the quality of the prediction of the dependent variable. It corresponds to the squared correlation between the predicted and the actual values of the dependent variable. It can also be interpreted as the proportion of the variance of the dependent variable explained by the independent variables. When the independent variables (used for predicting the dependent variable) are pair wise orthogonal, the multiple correlation coefficient is equal to the sum of the squared coefficients of correlation between each independent variable and the dependent variable. This relation does not hold when the independent variables are not orthogonal. The significance of a multiple coefficient of correlation can be assessed with an F ratio. The magnitude of the multiple coefficient of correlation tends to overestimate the magnitude of the population correlation, but it is possible to correct for this overestimation.

Caution Strictly speaking we should refer to this coefficient as the squared multiple correlation coefficient, but current usage seems to ignore the adjective "squared," probably because mostly its squared value is considered.

Task Distinguish between correlation and regression.

Self Assessment

Fill in the blanks:

5. ..... correlation is used as a measure of the degree of association in situations where the nature of population, from which data are collected, is not known.

Notes

6. A ..... rank correlation implies that a high (low) rank of an individual according to one characteristic is accompanied by its high (low) rank according to the other.
7. The regression equations are useful for predicting the value of ..... variable for given value of the ..... variable.
8. When two or more individuals have the same rank, each individual is assigned a rank equal to the ..... of the ranks that would have been assigned to them in the event of there being slight differences in their values.

10.5 Summary

- Researchers sometimes put all the data together, as if they were one sample.
- There are two simple ways to approach these types of data.
- We can use the technique of correlation to test the statistical significance of the association.
- In other cases we use regression analysis to describe the relationship precisely by means of an equation that has predictive value.
- Straight-line (linear) relationships are particularly important because a straight line is a simple pattern that is quite common.
- The correlation measures the direction and strength of the linear relationship.

10.6 Keywords

Correlation: It is an analysis of covariation between two or more variables.

Correlation Coefficient: It is a numerical measure of the degree of association between two or more variables.

10.7 Review Questions

1. Obtain the two lines of regression from the following data and estimate the blood pressure when age is 50 years. Can we also estimate the blood pressure of a person aged 20 years on the basis of this regression equation? Discuss.

Age (X) (in years)	56	42	72	39	63	47	52	49	40	42	68	60
Blood Pressure (Y)	127	112	140	118	129	116	130	125	115	120	135	133

2. Show that the coefficient of correlation, r, is independent of change of origin and scale.
3. Prove that the coefficient of correlation lies between - 1 and + 1.
4. "If two variables are independent the correlation between them is zero, but the converse is not always true". Explain the meaning of this statement.
5. What is Spearman's rank correlation? What are the advantages of the coefficient of rank correlation over Karl Pearson's coefficient of correlation?

Answers: Self Assessment

1. 'Spearman's Rank Correlation
2. degree, direction
3. probable error
4. extreme

Notes

5. Rank
6. positive
7. dependent, independent
8. mean

10.8 Further Readings

Books

Abrams, M.A., *Social Surveys and Social Action*, London: Heinemann, 1951.

Arthur, Maurice, *Philosophy of Scientific Investigation* , Baltimore: John Hopkins University Press, 1943.

R.S. Bhardwaj, *Business Statistics*, Excel Books, New Delhi, 2008.

S.N. Murthy and U. Bhojanna, *Business Research Methods*, Excel Books, 2007.

Online links

www.indiastudychannel.com

www.scribd.com/doc

www.soas.ac.uk

www.web-source.net

Unit 11: Multiple Regression and Correlation Analysis

Notes

CONTENTS

Objectives
Introduction
11.1 Regression Analysis
11.1.1 Simple Regression
11.2 Meaning of Multiple Regressions
11.3 Coefficient of Determination (γ^2)
11.3.1 Linear Multiple Regression Analysis
11.3.2 Logistic Regression Analysis
11.4 Coefficient of Multiple Determinations
11.5 Summary
11.6 Keywords
11.7 Review Questions
11.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss the Regression analysis;
- Explain the meaning of multiple regressions;
- Describe the coefficient of determination;
- Identify the coefficient of multiple determinations.

Introduction

As you develop Cause & Effect diagrams based on data, you may wish to examine the degree of correlation between variables. A statistical measurement of correlation can be calculated using the least squares method to quantify the strength of the relationship between two variables. The output of that calculation is the **Correlation Coefficient, or (r)**, which ranges between –1 and 1. A value of 1 indicates perfect positive correlation – as one variable increases, the second increases in a linear fashion. Likewise, a value of –1 indicates perfect negative correlation – as one variable increases, the second decreases. A value of zero indicates zero correlation.

Before calculating the Correlation Coefficient, the first step is to construct a scatter diagram. Most spreadsheets, including Excel, can handle this task. In this case, the process improvement team is analyzing door closing efforts to understand what the causes could be. The Y-axis represents the width of the gap between the sealing flange of a car door and the sealing flange on the body – a measure of how tight the door is set to the body. The fishbone diagram indicated that variability in the seal gap could be a cause of variability in door closing efforts.

Notes

Notes It is important to note that Correlation is not Causation - two variables can be very strongly correlated, but both can be caused by a third variable.

Example: Consider two variables: (1) how much my grass grows per week, and (2) the average depth of the local reservoir. Both variables could be highly correlated because both are dependent upon a third variable – how much it rains.

11.1 Regression Analysis

If the coefficient of correlation calculated for bivariate data (X_i, Y_i), $i = 1, 2, ... n$, is reasonably high and a cause and effect type of relation is also believed to be existing between them, the next logical step is to obtain a functional relation between these variables. This functional relation is known as regression equation in statistics. Since the coefficient of correlation is measure of the degree of linear association of the variables, we shall discuss only linear regression equation. This does not, however, imply the non-existence of non-linear regression equations.

The regression equations are useful for predicting the value of dependent variable for given value of the independent variable. As pointed out earlier, the nature of a regression equation is different from the nature of a mathematical equation, e.g., if $Y = 10 + 2X$ is a mathematical equation then it implies that Y is exactly equal to 20 when $X = 5$. However, if $Y = 10 + 2X$ is a regression equation, then $Y = 20$ is an average value of Y when $X = 5$.

The term regression was first introduced by Sir Francis Galton in 1877. In his study of the relationship between heights of fathers and sons, he found that tall fathers were likely to have tall sons and vice-versa. However, the mean height of sons of tall fathers was lower than the mean height of their fathers and the mean height of sons of short fathers was higher than the mean height of their fathers. In this way, a tendency of the human race to regress or to return to a normal height was observed. Sir Francis Galton referred this tendency of returning to the mean height of all men as regression in his research paper, "Regression towards mediocrity in hereditary stature". The term 'Regression', originated in this particular context, is now used in various fields of study, even though there may be no existence of any regressive tendency.

11.1.1 Simple Regression

For a bivariate data (X_i, Y_i), $i = 1, 2, ... n$, we can have either X or Y as independent variable. If X is independent variable then we can estimate the average values of Y for a given value of X. The relation used for such estimation is called regression of Y on X. If on the other hand Y is used for estimating the average values of X, the relation will be called regression of X on Y. For a bivariate data, there will always be two lines of regression. It will be shown later that these two lines are different, i.e., one cannot be derived from the other by mere transfer of terms, because the derivation of each line is dependent on a different set of assumptions.

Line of Regression of Y on X

The general form of the line of regression of Y on X is $Y_{Ci} = a + bX_i$, where Y_{Ci} denotes the average or predicted or calculated value of Y for a given value of $X = X_i$. This line has two constants, a and b. The constant a is defined as the average value of Y when $X = 0$. Geometrically, it is the intercept of the line on Y-axis. Further, the constant b, gives the average rate of change of Y per unit change in X, is known as the regression coefficient.

Notes

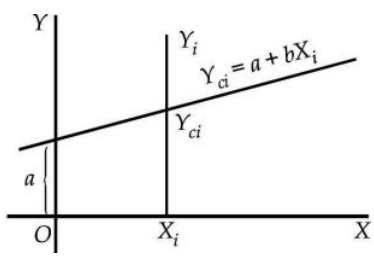
The above line is known if the values of a and b are known. These values are estimated from the observed data (X_i, Y_i), $i = 1, 2, ... n$.

Notes It is important to distinguish between YCi and Yi. Where as Yi is the observed value, YCi is a value calculated from the regression equation.

Using the regression $Y_{Ci} = a + bX_i$, we can obtain $Y_{C1}, Y_{C2}, ... Y_{Cn}$ corresponding to the X values $X_1, X_2, ... X_n$ respectively. The difference between the observed and calculated value for a particular value of X say X_i is called error in estimation of the i th observation on the assumption of a particular line of regression. There will be similar type of errors for all the n observations. We denote by $e_i = Y_i - Y_{Ci}$ $(i = 1, 2,.....n)$, the error in estimation of the i th observation. As is obvious from Figure 11.1, ei will be positive if the observed point lies above the line and will be negative if the observed point lies below the line. Therefore, in order to obtain a Figure of total error, ei's are squared and added. Let S denote the sum of squares of these errors,

i.e., $S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - Y_{Ci}\right)^2$.

Figure 11.1



The regression line can, alternatively, be written as a deviation of Y_i from Y_{Ci} i.e. $Y_i - Y_{Ci} = e_i$ or $Y_i = Y_{Ci} + e_i$ or $Y_i = a + bX_i + e_i$. The component $a + bXi$ is known as the deterministic component and ei is random component.

The value of S will be different for different lines of regression. A different line of regression means a different pair of constants a and b. Thus, S is a function of a and b. We want to find such values of a and b so that S is minimum. This method of finding the values of a and b is known as the Method of Least Squares.

Rewrite the above equation as $S = \Sigma(Yi - a - bX_i)^2$ $(\because YCi = a + bXi)$.

The necessary conditions for minima of S are

(i) $\frac{\partial S}{\partial a} = 0$ and (ii) $\frac{\partial S}{\partial b} = 0$, where $\frac{\partial S}{\partial a}$ and $\frac{\partial S}{\partial b}$ are the partial derivatives of S w.r.t. a and b respectively.

Now

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{n}\left(Y_i - a - bX_i\right) = 0$$

Or $$\sum_{i=1}^{n}\left(Y_i - a - bX_i\right) = \sum_{i=1}^{n} Y_i - na - b\sum_{i=1}^{n} X_i = 0$$

Notes

or $$\sum_{i=1}^{n} Y_i = na + b\sum_{i=1}^{n} X_i \quad .... (1)$$

Also, $$\frac{\partial S}{\partial b} = 2\sum_{i=1}^{n}\left(Y_i - a - bX_i\right)\left(-X_i\right) = 0$$

or $$-2\sum_{i=1}^{n}\left(X_iY_i - aX_i - bX_i^2\right) = \sum_{i=1}^{n}\left(X_iY_i - aX_i - bX_i^2\right) = 0$$

or $$\sum_{i=1}^{n} X_iY_i - a\sum_{i=1}^{n} X_i - b\sum_{i=1}^{n} X_i^2 = 0$$

or $$\sum_{i=1}^{n} X_iY_i = a\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X_i^2 \quad .... (2)$$

Equations (1) and (2) are a system of two simultaneous equations in two unknowns *a* and *b*, which can be solved for the values of these unknowns. These equations are also known as normal equations for the estimation of *a* and *b*. Substituting these values of *a* and *b* in the regression equation $Y_{Ci} = a + b_{Xi}$, we get the estimated line of regression of *Y* on *X*.

Expressions for the Estimation of *a* and *b*.

Dividing both sides of the equation (1) by *n*, we have

$$\frac{\sum Y_i}{n} = \frac{na}{n} + \frac{b\sum X_i}{n} \text{ or } \bar{Y} = a + b\bar{X} \quad .... (3)$$

This shows that the line of regression $Y_{Ci} = a + bX_i$ passes through the point $\left(\bar{X}, \bar{Y}\right)$.

From equation (3), we have $$a = \bar{Y} - b\bar{X} \quad .... (4)$$

Substituting this value of a in equation (2), we have

$$S X_i Y_i = \left(\bar{Y} - b\bar{X}\right)\Sigma X_i + b\Sigma X_i^2$$

$$= \bar{Y}\sum X_i - b\bar{X}\sum X_i + b\sum X_i^2 = n\bar{X}\bar{Y} - b.n\bar{X}^2 + b\sum X_i^2$$

or $$\sum X_iY_i - n\bar{X}\bar{Y} = b\left(\sum X_i^2 - n\bar{X}^2\right)$$

or $$b = \frac{\sum X_iY_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} \quad .... (5)$$

Also, $$\sum X_iY_i - n\bar{X}\bar{Y} = \sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$$

and $$\sum X_i^2 - n\bar{X}^2 = \sum\left(X_i - \bar{X}\right)^2$$

$\therefore$ $$b = \frac{\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum\left(X_i - \bar{X}\right)^2} \quad .... (6)$$

Notes

or $$b = \frac{\sum x_i y_i}{\sum x_i^2} \quad \text{.... (7)}$$

where x_i and y_i are deviations of values from their arithmetic mean.

Dividing numerator and denominator of equation (6) by n we have

$$b = \frac{\frac{1}{n}\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n}\sum(X_i - \bar{X})^2} = \frac{Cov(X,Y)}{\sigma_X^2} \quad \text{.... (8)}$$

The expression for b, which is convenient for use in computational work, can be written from equation (5) is given below:

$$b = \frac{\sum X_i Y_i - n\frac{\sum X_i}{n}\cdot\frac{\sum Y_i}{n}}{\sum X_i^2 - n\left(\frac{\sum X_i}{n}\right)^2} = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

Multiplying numerator and denominator by n, we have

$$b = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n\sum X_i^2 - (\sum X_i)^2} \quad \text{.... (9)}$$

To write the shortcut formula for b, we shall show that it is independent of change of origin but not of change of scale.

As in case of coefficient of correlation we define

$$u_i = \frac{X_i - A}{h} \quad \text{and} \quad v_i = \frac{Y_i - B}{k}$$

or $$X_i = A + hu_i \quad \text{and} \quad Y_i = B + kv_i$$

$\therefore$ $$\bar{X} = A + h\bar{u} \quad \text{and} \quad \bar{Y} = B + k\bar{v}$$

also $$(X_i - \bar{X}) = h(u_i - \bar{u}) \quad \text{and} \quad Y_i - \bar{Y} = k(v_i - \bar{v})$$

Substituting these values in equation (6), we have

$$b = \frac{hk\sum(u_i - \bar{u})(v_i - \bar{v})}{h^2\sum(u_i - \bar{u})^2} = \frac{k\sum(u_i - \bar{u})(v_i - \bar{v})}{h\sum(u_i - \bar{u})^2}$$

$$= \frac{k}{h}\left[\frac{n\sum u_i v_i - (\sum u_i)(\sum v_i)}{n\sum u_i^2 - (\sum u_i)^2}\right] \quad \text{.... (10)}$$

(***Note:*** if $h = k$ they will cancel each other)

Notes

Consider equation (8), $b = \frac{Cov(X,Y)}{\sigma_X^2}$

Writing $Cov(X, Y) = r \times \sigma_X \sigma_Y$, we have $b = \frac{r \cdot \sigma_X \sigma_Y}{\sigma_X^2} = r \cdot \frac{\sigma_Y}{\sigma_X}$

The line of regression of Y on X, i.e $Y_{Ci} = a + bX_i$ can also be written as

or $$Y_{Ci} = \bar{Y} - b\bar{X} + bX_i \text{ or } Y_{Ci} - \bar{Y} = b(X_i - \bar{X}) \quad \text{.... (11)}$$

or $$(Y_{Ci} - \bar{Y}) = r \cdot \frac{\sigma_Y}{\sigma_X}(X_i - \bar{X}) \quad \text{.... (12)}$$

Line of Regression of X on Y

The general form of the line of regression of X on Y is $X_{Ci} = c + dY_i$, where X_{Ci} denotes the predicted or calculated or estimated value of X for a given value of $Y = Y_i$ and c and d are constants. d is known as the regression coefficient of regression of X on Y.

In this case, we have to calculate the value of c and d so that

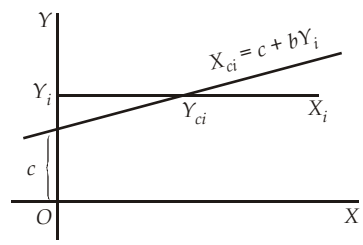
$$S' = (X_i - X_{Ci})^2 \text{ is minimised.}$$

As in the previous section, the normal equations for the estimation of c and d are

$$\Sigma X_i = nc + d\Sigma Y_i \quad \text{.... (13)}$$

and $$\Sigma X_i Y_i = c\Sigma Y_i + d\Sigma Y_i^2 \quad \text{.... (14)}$$

Figure 11.2



Dividing both sides of equation (13) by n, we have $\bar{X} = c + d\bar{Y}$.

This shows that the line of regression also passes through the point $(\bar{X}, \bar{Y})$. Since both the lines of regression passes through the point $(\bar{X}, \bar{Y})$, therefore $(\bar{X}, \bar{Y})$ is their point of intersection as shown in Figure 11.3.

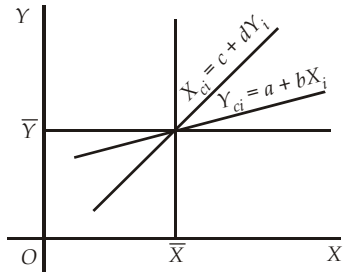
We can write $$c = \bar{X} - d\bar{Y} \quad \text{.... (15)}$$

As before, the various expressions for d can be directly written, as given below.

$$d = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum Y_i^2 - n\bar{Y}^2} \quad \text{.... (16)}$$

Notes

Figure 11.3



or $$d = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2} \quad .... (17)$$

or $$d = \frac{\sum x_i y_i}{\sum y_i^2} \quad .... (18)$$

$$= \frac{\frac{1}{n}\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n}\sum(Y_i - \bar{Y})^2} = \frac{Cov(X,Y)}{\sigma_Y^2} \quad .... (19)$$

Also $$d = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n\sum Y_i^2 - (\sum Y_i)^2} \quad .... (20)$$

This expression is useful for calculating the value of d. Another short-cut formula for the calculation of d is given by

$$d = \frac{h}{k}\left[\frac{n\sum u_i v_i - (\sum u_i)(\sum v_i)}{n\sum v_i^2 - (\sum v_i)^2}\right] \quad .... (21)$$

where $u_i = \frac{X_i - A}{h}$ and $v_i = \frac{Y_i - B}{k}$

Consider equation (19)

$$d = \frac{Cov(X,Y)}{\sigma_Y^2} = \frac{r\sigma_X\sigma_Y}{\sigma_Y^2} = r \cdot \frac{\sigma_X}{\sigma_Y} \quad .... (22)$$

Substituting the value of c from equation (15) into line of regression of X on Y we have

$$X_{Ci} = \bar{X} - d\bar{Y} + dY_i \text{ or } (X_{Ci} - \bar{X}) = d(Y_i - \bar{Y}) \quad .... (23)$$

or $$(X_{Ci} - \bar{X}) = r \cdot \frac{\sigma_X}{\sigma_Y}(Y_i - \bar{Y}) \quad .... (24)$$

Remarks: It should be noted here that the two lines of regression are different because these have been obtained in entirely two different ways. In case of regression of *Y* on *X*, it is assumed

Notes

that the values of X are given and the values of Y are estimated by minimising $S(Y_i - Y_{Ci})^2$ while in case of regression of X on Y, the values of Y are assumed to be given and the values of X are estimated by minimising $S(X_i - X_{Ci})^2$. Since these two lines have been estimated on the basis of different assumptions, they are not reversible, i.e., it is not possible to obtain one line from the other by mere transfer of terms. There is, however, one situation when these two lines will coincide. From the study of correlation we may recall that when $r = \pm 1$, there is perfect correlation between the variables and all the points lie on a straight line. Therefore, both the lines of regression coincide and hence they are also reversible in this case. By substituting $r = \pm 1$ in equation (12) or (24) it can be shown that the lines of regression in both the cases become

$$\left(\frac{Y_i - \bar{Y}}{\sigma_Y}\right) = \pm\left(\frac{X_i - \bar{X}}{\sigma_X}\right)$$

Further when $r = 0$, equation (12) becomes $Y_{Ci} = \bar{Y}$ and equation (24) becomes $X_{Ci} = \bar{X}$. These are the equations of lines parallel to X-axis and Y-axis respectively. These lines also intersect at the point $(\bar{X}, \bar{Y})$ and are mutually perpendicular at this point.

Correlation Coefficient and the Two Regression Coefficients

Since $b = r \cdot \frac{\sigma_Y}{\sigma_X}$ and $d = r \cdot \frac{\sigma_X}{\sigma_Y}$, we have

$b \cdot d = r\frac{\sigma_Y}{\sigma_X} \cdot r\frac{\sigma_X}{\sigma_Y} = r^2$ or $r = \sqrt{b \cdot d}$. This shows that correlation coefficient is the geometric mean of the two regression coefficients.

Remarks: The following points should be kept in mind about the coefficient of correlation and the regression coefficients:

1. Since $r = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$, $b = \frac{Cov(X,Y)}{\sigma_X^2}$ and $d = \frac{Cov(X,Y)}{\sigma_Y^2}$, therefore the sign of r, b and d will always be same and this will depend upon the sign of $Cov(X, Y)$.
2. Since $bd = r^2$ and $0 \le r^2 \le 1$, therefore either both b and d are less than unity or if one of them is greater than unity, the other must be less than unity such that $0 \le b \times d \le 1$ is always true.

Example: Obtain the two regression equations and find correlation coefficient between X and Y from the following data:

X	10	9	7	8	11
Y	6	3	2	4	5

Solution:

Calculation Table

X	Y	XY	X^2	Y^2
10	6	60	100	36
9	3	27	81	9

Contd...

Notes

7	2	14	49	4
8	4	32	64	16
11	5	55	121	25
45	**20**	**188**	**415**	**90**

(a) Regression of Y on X

$$b = \frac{n\sum XY - \left(\sum X\right)\left(\sum Y\right)}{n\sum X^2 - \left(\sum X\right)^2} = \frac{5 \times 188 - 45 \times 20}{5 \times 415 - (45)^2} = 0.8$$

Also, $\bar{X} = \frac{45}{5} = 9$ and $\bar{Y} = \frac{20}{5} = 4$

Now $a = \bar{Y} - b\bar{X} = 4 - 0.8 \times 9 = -3.2$

$\therefore$ Regression of Y on X is $Y_C = -3.2 + 0.8X$

(b) Regression of X on Y

$$d = \frac{n\sum XY - \left(\sum X\right)\left(\sum Y\right)}{n\sum Y^2 - \left(\sum Y\right)^2} = \frac{5 \times 188 - 45 \times 20}{5 \times 90 - (20)^2} = 0.8$$

Also, $c = \bar{X} - d\bar{Y}$ $9 - 0.8 \times 4 = 5.8$

$\therefore$ The regression of X on Y is $X_C = 5.8 + 0.8Y$

(c) Coefficient of correlation $r = \sqrt{b \cdot d} = \sqrt{0.8 \times 0.8} = 0.8$

Example: From the data given below, find:

1. The two regression equations.
2. The coefficient of correlation between marks in economics and statistics.
3. The most likely marks in statistics when marks in economics are 30.

Marks in Eco.	25	28	35	32	31	36	29	38	34	32
Marks in Stat.	43	46	49	41	36	32	31	30	33	39

Solution:

Calculation Table

Marks in Eco. (X)	**Marks in Stat. (Y)**	**u = X – 31**	**v = Y – 41**	**uv**	**u^2**	**v^2**
25	43	– 6	2	– 12	36	4
28	46	– 3	5	– 15	9	25
35	49	4	8	32	16	64
32	41	1	0	0	1	0
31	36	0	– 5	0	0	25

Contd...

Notes

36	32	5	– 9	– 45	25	81
29	31	– 2	– 10	20	4	100
38	30	7	– 11	– 77	49	121
34	33	3	– 8	– 24	9	64
32	39	1	– 2	– 2	1	4
Total		**10**	**– 30**	**– 123**	**150**	**488**

From the table, we have

$$\bar{X} = 31 + \frac{10}{10} = 32 \text{ and } \bar{Y} = 41 - \frac{30}{10} = 38$$

1. The lines of regression

   (a) Regression of Y on X

$$b = \frac{n\sum uv - \left(\sum u\right)\left(\sum v\right)}{n\sum u^2 - \left(\sum u\right)^2} = \frac{-1230 + 300}{1500 - 100} = -0.66$$

$$a = \bar{Y} - b\bar{X} = 38 + 0.66 \times 32 = 59.26$$

   ∴ Regression equation is

$$Y_C = 59.26 - 0.66X$$

   (b) Regression of X on Y

$$d = \frac{n\sum uv - \left(\sum u\right)\left(\sum v\right)}{n\sum v^2 - \left(\sum v\right)^2} = \frac{-1230 + 300}{4880 - 900} = -0.23$$

$$c = \bar{X} - d\bar{Y} = 32 + 0.23 \times 38 = 40.88$$

   ∴ Regression equation is

$$X_C = 40.88 - 0.23Y$$

2. Coefficient of correlation

$$r = \sqrt{b \cdot d} = -\sqrt{-0.66 \times -0.23} = -0.39$$

Note that r, b and d are of same sign.

3. Since we have to estimate marks in statistics denoted by Y, therefore, regression of Y on X will be used. The most likely marks in statistics when marks in economics are 30, is given by

$$Y_C = 59.26 - 0.66 \times 30 = 39.33$$

Self Assessment

Fill in the blanks:

1. The regression line can, alternatively, be written as a deviation of Y_i from Y_{Ci} i.e. $Y_i - Y_{Ci}$ = ……

2. A different ……..means a different pair of constants a and b.

11.2 Meaning of Multiple Regressions

Notes

Multiple regressions are a statistical technique that allows us to predict someone's score on one variable on the basis of their scores on several other variables. An example might help. Suppose we were interested in predicting how much an individual enjoys their job. Variables such as salary, extent of academic qualifications, age, sex, number of years in full-time employment and socioeconomic status might all contribute towards job satisfaction. If we collected data on all of these variables, perhaps by surveying a few hundred members of the public, we would be able to see how many and which of these variables gave rise to the most accurate prediction of job satisfaction. We might find that job satisfaction is most accurately predicted by type of occupation, salary and years in full-time employment, with the other variables not helping us to predict job satisfaction.

When using multiple regressions in psychology, many researchers use the term "independent variables" to identify those variables that they think will influence some other "dependent variable". We prefer to use the term "predictor variables" for those variables that may be useful in predicting the scores on another variable that we call the "criterion variable". Thus, in our example above, type of occupation, salary and years in full-time employment would emerge as significant predictor variables, which allow us to estimate the criterion variable – how satisfied someone is likely to be with their job. As we have pointed out before, human behaviour is inherently noisy and therefore it is not possible to produce totally accurate predictions, but multiple regressions allow us to identify a set of predictor variables which together provide a useful estimate of a participant's likely score on a criterion variable.

In the case of simple linear regression, one variable, say, X_1 is affected by a linear combination of another variable X_2 (we shall use X_1 and X_2 instead of Y and X used earlier). However, if X_1 is affected by a linear combination of more than one variable, the regression is termed as a multiple linear regression.

Let there be k variables X_1, X_2 ..... X_k, where one of these, say X_j, is affected by the remaining k – 1 variables. We write the typical regression equation as

$$X_{jc} = a_{j\times 1,\,2,\,....\,j-1,\,j+1,\,....\,k} + b_{j\,1.2,3,\,....\,j-1,\,j+1,\,....k}X_1 + b_{j\,2.1,\,3,\,....\,j-1,\,j+1,\,....k}X_2 + ...(j = 1, 2, ... k).$$

Here $a_{j.1,2,\,....}$, $b_{j1.2,\,3,\,....}$ ..... etc. are constants. The constant $aj.1,2,$ .... is interpreted as the value of X_j when X_2, X_3, ... X_{j-1}, X_{j+1} ..... X_k are all equal to zero. Further, $b_{j1.2,3,\,....\,j-1,\,j+1,\,....k}$, $b_{j2.1,3,\,....\,j-1,\,j+1,\,....k}$ etc., are (k – 1) partial regression coefficients of regression of X_j on X_1, X_2 ..... X_{j-1}, X_{j+1} ..... X_k.

For simplicity, we shall consider three variables $X1$, $X2$ and $X3$. The three possible regression equations can be written as

$$X_{1c} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3 \quad .... (1)$$

$$X_{2c} = a_{2.13} + b_{21.3}X_1 + b_{23.1}X_3 \quad .... (2)$$

$$X_{3c} = a_{3.12} + b_{31.2}X_1 + b_{32.1}X_2 \quad .... (3)$$

Given n observations on X_1, X_2 and X_3, we want to find such values of the constants of the regression equation so that $\sum_{i=1}^{n}\left(X_{ij} - X_{ijc}\right)^2$, j = 1, 2, 3, is minimised.

Caution For convenience, we shall use regression equations expressed in terms of deviations of variables from their respective means.

Notes

Equation (1), on taking sum and dividing by n, can be written as

$$\frac{\sum X_{1c}}{n} = a_{1.23} + b_{12.3}\frac{\sum X_2}{n} + b_{13.2}\frac{\sum X_3}{n} \quad \text{or} \quad \bar{X}_1 = a_{1.23} + b_{12.3}\bar{X}_2 + b_{13.2}\bar{X}_3 \qquad \text{.... (4)}$$

Notes $\Sigma X_1 = \Sigma X_{1c}$.

Subtracting (4) from (1), we have

$$X_{1c} - \bar{X}_1 = b_{12.3}\left(X_2 - \bar{X}_2\right) + b_{13.2}\left(X_3 - \bar{X}_3\right) \text{ or } x_{1c} = b_{12.3}x_2 + b_{13.2}x_3 \qquad \text{.... (5)}$$

where $X_{1c} - \bar{X}_1 = x_{1c}$, $X_2 - \bar{X}_2 = x_2$ and $X_3 - \bar{X}_3 = x_3$.

Similarly, we can write equations (2) and (3) as

$$x_{2c} = b_{21.3}x_1 + b_{23.1}x_3 \qquad \text{.... (6)}$$

and $x_{3c} = b_{31.2}x_1 + b_{32.1}x_2$, respectively. .... (7)

Notes The subscript of the coefficients preceding the dot are termed as primary subscripts while those appearing after it are termed as secondary subscripts. The number of secondary subscripts gives the order of the regression coefficient, e.g., $b_{12.3}$ is regression coefficient of order one, etc.

Least Square Estimates of Regression Coefficients

Let us first estimate the coefficients of regression equation (5). Given n observations on each of the three variables X_1, X_2 and X_3, we have to find the values of the constants $b_{12.3}$ and $b_{13.2}X_3$ so that is minimised. Using method of least squares, the normal equations can be written as

$$\sum x_1x_2 = b_{12.3}\sum x_2^2 + b_{13.2}\sum x_2x_3 \qquad \text{.... (8)}$$

$$\sum x_1x_3 = b_{12.3}\sum x_2x_3 + b_{13.2}\sum x_3^2 \qquad \text{.... (9)}$$

Solving the above equations simultaneously, we get

$$b_{12.3} = \frac{\left(\sum x_1x_2\right)\left(\sum x_3^2\right) - \left(\sum x_1x_3\right)\left(\sum x_2x_3\right)}{\left(\sum x_2^2\right)\left(\sum x_3^2\right) - \left(\sum x_2x_3\right)^2} \qquad \text{.... (10)}$$

$$b_{13.2} = \frac{\left(\sum x_1x_3\right)\left(\sum x_2^2\right) - \left(\sum x_1x_2\right)\left(\sum x_2x_3\right)}{\left(\sum x_2^2\right)\left(\sum x_3^2\right) - \left(\sum x_2x_3\right)^2} \qquad \text{.... (11)}$$

Using equation (4), we can find $a_{1.23} = \bar{X}_1 - b_{12.3}\bar{X}_2 - b_{13.2}\bar{X}_3$.

Note:

Notes

1. Various sums of squares and sums of products of deviations, used above, can be computed using the formula $\sum x_p x_q = \sum X_p X_q - \frac{\left(\sum X_p\right)\left(\sum X_q\right)}{n}$. For example, put p = 1 and q = 2 in the formula to obtain $\Sigma X_1 X_2$ and put $p = q = 2$, to obtain $\sum x_2^2$, etc.

2. The fact that a regression coefficient is independent of change of origin can also be utilised to further simplify the computational work.

3. The regression coefficients of equations (2) and (3) can be written by symmetry as given below:

$$b_{21.3} = \frac{\left(\sum x_2 x_1\right)\left(\sum x_3^2\right) - \left(\sum x_2 x_3\right)\left(\sum x_1 x_3\right)}{\left(\sum x_1^2\right)\left(\sum x_3^2\right) - \left(\sum x_1 x_3\right)^2}$$

$$b_{23.1} = \frac{\left(\sum x_2 x_3\right)\left(\sum x_1^2\right) - \left(\sum x_2 x_1\right)\left(\sum x_1 x_3\right)}{\left(\sum x_1^2\right)\left(\sum x_3^2\right) - \left(\sum x_1 x_3\right)^2}$$

Further, $b_{31.2} = b_{13.2}$ and $b_{32.1} = b_{23.1}$ and the expressions for the constant terms are $a_{2.13} = \bar{X}_2 - b_{21.3}\bar{X}_1 - b_{23.1}\bar{X}_3$ and $a_{3.12} = \bar{X}_3 - b_{31.2}\bar{X}_1 - b_{32.1}\bar{X}_2$ respectively.

11.3 Coefficient of Determination (γ^2)

When = 1; or -1; or 0, the interpretation of does not pose any problem. When = 1; or –1, all the points lie on straight line in a graph showing a perfect positive or negative correlation. When the points are extremely scattered on a graph, then it becomes evident that there is almost no relationship between the two variables. However, when it comes to other values of, we have to be careful in its interpretation. Suppose we get a correlation of = 0.9, we may say that = 0.9 is 'twice as good' or 'twice as strong' as a correlation of = 0.45. It may be noted that this comparison is wrong. The strength of is judged by coefficient of determination, for = 0.9, = 0.81. We multiply it by 100, thus getting 81 per cent. Thus suggest that when = 0.9 then we can say that 81 per cent of the total variation in the Y series can be attributed to the relationship with X.

11.3.1 Linear Multiple Regression Analysis

Multiple regressions is the most commonly utilized multivariate technique. It examines the relationship between a single metric dependent variable and two or more metric independent variables. The technique relies upon determining the linear relationship with the lowest sum of squared variances; therefore, assumptions of normality, linearity, and equal variance are carefully observed. The beta coefficients (weights) are the marginal impacts of each variable, and the size of the weight can be interpreted directly. Multiple regression is often used as a forecasting tool.

11.3.2 Logistic Regression Analysis

Sometimes referred to as "choice models," this technique is a variation of multiple regressions that allows for the prediction of an event. It is allowable to utilize non-metric (typically binary)

Notes

dependent variables, as the objective is to arrive at a probabilistic assessment of a binary choice. The independent variables can be either discrete or continuous. A contingency table is produced, which shows the classification of observations as to whether the observed and predicted events match. The sum of events that were predicted to occur which actually did occur and the events that were predicted not to occur which actually did not occur, divided by the total number of events, is a measure of the effectiveness of the model. This tool helps predict the choices consumers might make when presented with alternatives.

11.4 Coefficient of Multiple Determinations

In statistics, the **coefficient of determination R^2** is used in the context of statistical models whose main purpose is the prediction of future outcomes on the basis of other related information. It is the proportion of variability in a data set that is accounted for by the statistical model. It provides a measure of how well future outcomes are likely to be predicted by the model.

There are several different definitions of R^2 which are only sometimes equivalent. One class of such cases includes that of linear regression. In this case, if an intercept is included then R^2 is simply the square of the sample correlation coefficient between the outcomes and their predicted values, or in the case of simple linear regression, between the outcomes and the values of the single regress or being used for prediction. In such cases, the coefficient of determination ranges from 0 to 1. Important cases where the computational definition of R^2 can yield negative values, depending on the definition used, arise where the predictions which are being compared to the corresponding outcomes have not been derived from a model-fitting procedure using those data, and where linear regression is conducted without including an intercept. Additionally, negative values of R^2 may occur when fitting non-linear trends to data. In these instances, the mean of the data provides a fit to the data that is superior to that of the trend under this goodness of fit analysis.

In multiple regression analysis, the proportion of the variation in Y explained by the regression, which can be calculated as SSexplained/SStotal . In other words this is the proportion of variation in the criterion variable that is accounted for by the co-variations in the predictor (independent) variable. The coefficient of determination of a multiple linear regression model is the quotient of the variances of the fitted values and observed values of the dependent variable. If we denote y_i as the observed values of the dependent variable, $\bar{y}$ as its mean, and $\hat{y}_i$ as the fitted value, then the coefficient of determination is:

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

Self Assessment

Fill in the blanks:

3. ……..analysis sometimes referred to as choice models.
4. In statistics, the …… R^2 is used in the context of statistical models whose main purpose is the prediction of future outcomes on the basis of other related information.

11.5 Summary

- If the coefficient of correlation calculated for bivariate data (X_i, Y_i), i = 1, 2, ... n, is reasonably high and a cause and effect type of relation is also believed to be existing between them, the next logical step is to obtain a functional relation between these variables.

Notes

- The general form of the line of regression of Y on X is $YCi = a + bXi$, where YCi denotes the average or predicted or calculated value of Y for a given value of $X = Xi$.
- Multiple regressions are a statistical technique that allows us to predict someone's score on one variable on the basis of their scores on several other variables. An example might help.
- There are several different definitions of R^2 which are only sometimes equivalent. One class of such cases includes that of linear regression.
- The least-squares regression line is the line that makes the sum of the squares of the vertical distances of the data points from the line as small as possible.
- Non-parametric regression analysis traces the dependence of a response variable on one or several predictors without specifying in advance the function that relates the response to the predictors.

11.6 Keywords

Coefficient of determination:In statistics, the **coefficient of determination** R^2 is used in the context of statistical models whose main purpose is the prediction of future outcomes on the basis of other related information.

Regression Equation: If the coefficient of correlation calculated for bivariate data (Xi, Yi), i = 1, 2, ... n, is reasonably high and a cause and effect type of relation is also believed to be existing between them, the next logical step is to obtain a functional relation between these variables. This functional relation is known as regression equation in statistics.

11.7 Review Questions

1. Distinguish between correlation and regression. Discuss least square method of fitting regression.
2. What do you understand by linear regression? Why there are two lines of regression? Under what condition(s) can there be only one line?
3. What do you think as the reason behind the two lines of regression being different?
4. For a bivariate data, which variable can we have as independent? Why?
5. What can you conclude on the basis of the fact that the correlation between body weight and annual income were high and positive?

Answers: Self Assessment

1. *ei*
2. Line of regression
3. Logistic regression
4. Coefficient of determination

Notes

11.8 Further Readings

Books

Abrams, M.A., *Social Surveys and Social Action*, London: Heinemann, 1951.

Arthur, Maurice, *Philosophy of Scientific Investigation* , Baltimore: John Hopkins University Press, 1943.

R.S. Bhardwaj, *Business Statistics*, Excel Books, New Delhi, 2008.

S.N. Murthy and U. Bhojanna, *Business Research Methods*, Excel Books, 2007.

Online links

www.indiastudychannel.com

www.scribd.com/doc

www.soas.ac.uk

www.web-source.net

Notes

Unit 12: Hypothesis Testing

CONTENTS

Objectives
Introduction
12.1 Meaning of Hypothesis
12.2 Statistical Testing Procedure
12.2.1 Formulate the Hypothesis
12.2.2 Statistical Significance Level
12.2.3 One-tailed and Two-tailed Tests
12.2.4 Degree of Freedom
12.2.5 Select Test Criteria
12.2.6 Compute
12.2.7 Make Decisions
12.3 Errors in Hypothesis Testing
12.4 Types of Tests
12.4.1 Parametric Test
12.4.2 Non-parametric Test
12.5 P-values
12.6 Summary
12.7 Keywords
12.8 Review Questions
12.9 Further Readings

Objectives

After studying this unit, you will be able to:

- Identify the steps involved in hypothesis testing;
- Explain the statistical testing procedure;
- Discuss the errors in hypothesis testing;
- Explain the types of tests.

Introduction

A statistical hypothesis test is a method of making statistical decisions using experimental data. In statistics, a result is called statistically significant if it is unlikely to have occurred by chance. The phrase "test of significance" was coined by Ronald Fisher: "Critical tests of this kind may be called tests of significance, and when such tests are available we may discover whether a second sample is or is not significantly different from the first."

Notes

Hypothesis testing is sometimes called confirmatory data analysis, in contrast to exploratory data analysis. In frequency probability, these decisions are almost always made using null-hypothesis tests; that is, ones that answer the question. Assuming that the null hypothesis is true, what is the probability of observing a value for the test statistic that is at least as extreme as the value that was actually observed? One use of hypothesis testing is deciding whether experimental results contain enough information to cast doubt on conventional wisdom.

12.1 Meaning of Hypothesis

A hypothesis is a tentative proposition relating to certain phenomenon, which the researcher wants to verify when required.

If the researcher wants to infer something about the total population from which the sample was taken, statistical methods are used to make inference. We may say that, while a hypothesis is useful, it is not always necessary. Many a time, the researcher is interested in collecting and analysing the data indicating the main characteristics without a hypothesis. Also, a hypothesis may be rejected but can never be accepted except tentatively. Further evidence may prove it wrong. It is wrong to conclude that since hypothesis was not rejected it can be accepted as valid.

What is a Null Hypothesis?

A null hypothesis is a statement about the population, whose credibility or validity the researcher wants to assess based on the sample.

A null hypothesis is formulated specifically to test for possible rejection or nullification. Hence the name 'null hypothesis'. Null hypothesis always states "no difference". It is this null hypothesis that is tested by the researcher.

12.2 Statistical Testing Procedure

1. Formulate the null hypothesis, with H_0 and HA, the alternate hypothesis.

   According to the given problem, H_0 represents the value of some parameter of population.
2. Select on appropriate test assuming H_0 to be true.
3. Calculate the value.
4. Select the level of significance other at 1% or 5%.
5. Find the critical region.
6. If the calculated value lies within the critical region, then reject H_o.
7. State the conclusion in writing.

12.2.1 Formulate the Hypothesis

The normal approach is to set two hypotheses instead of one, in such a way, that if one hypothesis is true, the other is false. Alternatively, if one hypothesis is false or rejected, then the other is true or accepted. These two hypotheses are:

(1) Null hypothesis

(2) Alternate hypothesis

Let us assume that the mean of the population is μo and the mean of the sample is x. Since we have assumed that the population has a mean of μo, this is our null hypothesis. We write this as

Notes

Hoμ = μo, where Ho is the null hypothesis. Alternate hypothesis is HA=μ. The rejection of null hypothesis will show that the mean of the population is not μ o. This implies that alternate hypothesis is accepted.

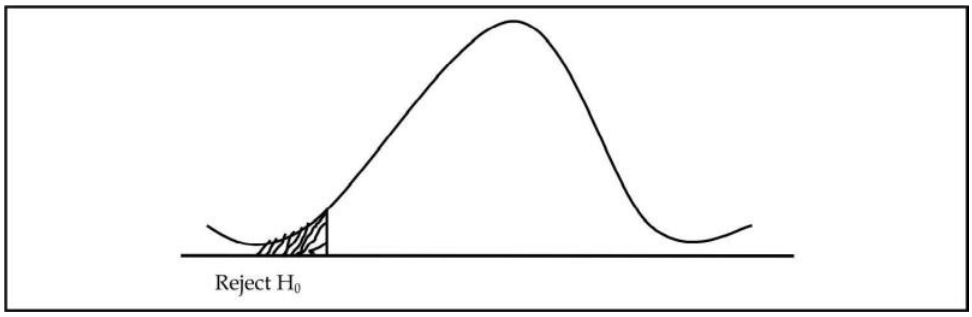
12.2.2 Statistical Significance Level

Having formulated the hypothesis, the next step is its validity at a certain level of significance. The confidence with which a null hypothesis is accepted or rejected depends upon the significance level. A significance level of say 5% means that the risk of making a wrong decision is 5%. The researcher is likely to be wrong in accepting false hypothesis or rejecting a true hypothesis by 5 out of 100 occasions. A significance level of say 1% means, that the researcher is running the risk of being wrong in accepting or rejecting the hypothesis is one of every 100 occasions. Therefore, a 1% significance level provides greater confidence to the decision than 5% significance level.

There are two type of tests.

12.2.3 One-tailed and Two-tailed Tests

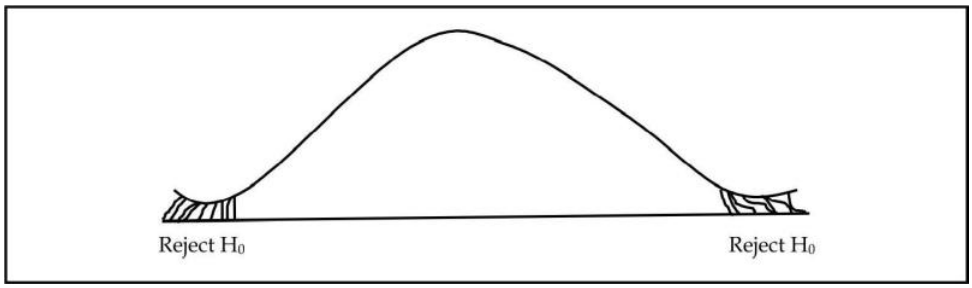
A hypothesis test may be one-tailed or two-tailed. In one-tailed test the test-statistic for rejection of null hypothesis falls only in one-tailed of sampling distribution curve.



Example:

- In a right side test, the critical region lies entirely in the right tail of the sample distribution. Whether the test is one-sided or two-sided – depends on alternate hypothesis.
- A tyre company claims that mean life of its new tyre is 15,000 km. Now the researcher formulates the hypothesis that tyre life is = 15,000 km.

A two-tailed test is one in which the test statistics leading to rejection of null hypothesis falls on both tails of the sampling distribution curve as shown.



When we should apply a hypothesis test that is one-tailed or two-tailed depends on the nature of the problem. One-tailed test is used when the researcher's interest is primarily on one side of the issue.

Notes

Example:

- "Is the current advertisement less effective than the proposed new advertisement"?
- A two-tailed test is appropriate, when the researcher has no reason to focus on one side of the issue.

Example: "Are the two markets – Mumbai and Delhi different to test market a product?"

- A product is manufactured by a semi-automatic machine. Now, assume that the same product is manufactured by the fully automatic machine. This will be two-sided test, because the null hypothesis is that "the two methods used for manufacturing the product do not differ significantly".

$\therefore \quad H_0 = \mu_1 = \mu_2$

Sign of alternate hypothesis	Type of test
=	Two-sided
<	One-sided to right
>	One-sided to left

12.2.4 Degree of Freedom

It tells the researcher the number of elements that can be chosen freely.

Example: a+b/2 =5. fix a=3, b has to be 7. Therefore, the degree of freedom is 1.

12.2.5 Select Test Criteria

If the hypothesis pertains to a larger sample (30 or more), the Z-test is used. When the sample is small (less than 30), the T-test is used.

12.2.6 Compute

Carry out computation.

12.2.7 Make Decisions

Accepting or rejecting of the null hypothesis depends on whether the computed value falls in the region of rejection at a given level of significance.

Task Discuss When Would You Prefer Two Tailed Test To One Tailed Test.

Self Assessment

Fill in the blanks:

1. A ……..is a tentative proposition relating to certain phenomenon, which the researcher wants to verify when required.
2. A ……..is a statement about the population, whose credibility or validity the researcher wants to assess based on the sample.

Notes

3. A ……test is one in which the test statistics leading to rejection of null hypothesis falls on both tails of the sampling distribution curve as shown.

4. ……tells the researcher the number of elements that can be chosen freely.

12.3 Errors in Hypothesis Testing

There are two types of errors:

1. Hypothesis is rejected when it is true.

2. Hypothesis is not rejected when it is false.

(1) is called Type 1 error (a), (2) is called Type 2 error (b). When a =0.10 it means that true hypothesis will be accepted in 90 out of 100 occasions. Thus, there is a risk of rejecting a true hypothesis in 10 out of every 100 occasions. To reduce the risk, use a = 0.01 which implies that we are prepared to take a 1% risk i.e., the probability of rejecting a true hypothesis is 1%. It is also possible that in hypothesis testing, we may commit Type 2 error (b) i.e., accepting a null hypothesis which is false.

Notes The only way to reduce Type 1 and Type 2 error is by increasing the sample size.

Example of Type 1 and Type 2 error:

Type 1 and Type 2 error is presented as follows. Suppose a marketing company has 2 distributors (retailers) with varying capabilities. On the basis of capabilities, the company has grouped them into two categories (1) Competent retailer (2) Incompetent retailer. Thus R 1 is a competent retailer and R2 is an incompetent retailer. The firm wishes to award a performance bonus (as a part of trade promotion) to encourage good retailership. Assume that two actions A1 and A2 would represent whether the bonus or trade incentive is given and not given. This is shown as follows:

Action	(R1) Competent retailer	(R2) Incompetent retailer
A 1 performance bonus is awarded	Correct decision	Incorrect decision error (β)
A 2 performance bonus is not awarded	Incorrect decision error (α)	Correct decision

When the firm has failed to reward a competent retailer, it has committed type-2 error. On the other hand, when it was rewarded to an incompetent retailer, it has committed type-1error.

12.4 Types of Tests

1. Parametric test.

2. Non-parametric test.

12.4.1 Parametric Test

(1) Parametric tests are more powerful. The data in this test is derived from interval and ratio measurement.

(2) In parametric tests, it is assumed that the data follows normal distributions. Examples of parametric tests are (a) Z-Test, (b) T-Test and (c) F-Test.

Notes

(3) Observations must be independent i.e., selection of any one item should not affect the chances of selecting any others be included in the sample.

Did u know? **What is univariate/bivariate data analysis?**

Univariate

If we wish to analyse one variable at a time, this is called univariate analysis. For example: Effect of sales on pricing. Here, price is an independent variable and sales is a dependent variable. Change the price and measure the sales.

Bivariate

The relationship of two variables at a time is examined by means of bi-variate data analysis.

If one is interested in a problem of detecting whether a parameter has either increased or decreased, a two-sided test is appropriate.

12.4.2 Non-parametric Test

Non-parametric tests are used to test the hypothesis with nominal and ordinal data.

(1) We do not make assumptions about the shape of population distribution.

(2) These are distribution-free tests.

(3) The hypothesis of non-parametric test is concerned with something other than the value of a population parameter.

(4) Easy to compute. There are certain situations particularly in marketing research, where the assumptions of parametric tests are not valid. For example: In a parametric test, we assume that data collected follows a normal distribution. In such cases, non-parametric tests are used. Examples of non-parametric tests are (a) Binomial test (b) Chi-Square test (c) Mann-Whitney U test (d) Sign test. A binominal test is used when the population has only two classes such as male, female; buyers, non-buyers, success, failure etc. All observations made about the population must fall into one of the two tests. The binomial test is used when the sample size is small.

Advantages

1. They are quick and easy to use.
2. When data are not very accurate, these tests produce fairly good results.

Disadvantages

Non-parametric test involves the greater risk of accepting a false hypothesis and thus committing a Type 2 error.

12.5 P-values

A p-value, sometimes called an uncertainty or probability coefficient, is based on properties of the sampling distribution. It is usually expressed as p less than some decimal, as in $p < .05$ or $p < .0006$, where the decimal is obtained by tweaking the significance setting of any statistical

Notes

procedure you run in SPSS. It is used in two ways: (1) as a *criterion level* where you, the researcher have arbitrarily decided in advance to use as the cutoff where you reject the null hypothesis, in which case, you would ordinarily say something like "setting p at p > .65 for one-tailed or two-tailed tests of significance allows some confidence that 65% of the time, rejecting the null hypothesis will not be in error"; and more commonly, (2) as a expression of *inference uncertainty* after you have run some test statistic regarding the strength of some association or relationship between your independent and dependent variables, in which case, you would say something like "the evidence suggests there is a statistically significant effect, however, p < .05 also suggests that 5% of the time, we should be uncertain about the significance of drawing any statistical inferences."

Task A study was conducted to measure the motivation level of each of the category of managers. Formulate a hypothesis, suggesting testing procedures to show that there is no relation between the category of managers and the level of motivation.

Self Assessment

Fill in the blanks:

5. To reduce the risk, use a = 0.01 which implies that we are prepared to take a ……..risk i.e., the probability of rejecting a true hypothesis is 1%.
6. In parametric tests, it is assumed that the data follows ……
7. ……test involves the greater risk of accepting a false hypothesis and thus committing a Type 2 error.
8. A …….sometimes called an uncertainty or probability coefficient, is based on properties of the sampling distribution.

12.6 Summary

- Hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true.
- The usual process of hypothesis testing consists of four steps.
- Formulate the null hypothesis and the alternative hypothesis.
- Identify a test statistic that can be used to assess the truth of the null hypothesis.
- Compute the P-value, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true.
- The smaller the -value, the stronger the evidence against the null hypothesis.
- Compare the -value to an acceptable significance value a.
- If , that the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid.

12.7 Keywords

Alternate Hypothesis: An alternative hypothesis is one that specifies that the null hypothesis is not true. The alternative hypothesis is false when the null hypothesis is true, and true when the null hypothesis is false.

Notes

Null Hypothesis: The null hypothesis is a hypothesis which the researcher tries to disprove, reject or nullify.

12.8 Review Questions

1. What hypothesis, test and procedure would you use when an automobile company has manufacturing facility at two different geographical locations? Each location manufactures two-wheelers of a different model. The customer wants to know if the mileage given by both the models is the same or not. Samples of 45 numbers may be taken for this purpose.
2. What hypothesis, test and procedure would you use when a company has 22 sales executives? They underwent a training programme. The test must evaluate whether the sales performance is unchanged or improved after the training programme.
3. What hypothesis, test and procedure would you use A company has three categories of managers:
   (a) With professional qualifications but without work experience.
   (b) With professional qualifications accompanied by work experience.
   (c) Without professional qualifications but with work experience.

Answers: Self Assessment

1. Hypothesis
2. Null hypothesis
3. Two-tailed
4. Degree of freedom
5. 1%
6. Normal distributions
7. Non-parametric
8. P-value

12.9 Further Readings

Books

Abrams, M.A, *Social Surveys and Social Action*, London: Heinemann, 1951.

Arthur, Maurice, *Philosophy of Scientific Investigation* , Baltimore: John Hopkins University Press, 1943.

R.S. Bhardwaj, *Business Statistics*, Excel Books, New Delhi, 2008.

S.N. Murthy and U. Bhojanna, *Business Research Methods*, Excel Books, 2007.

Online links

www.indiastudychannel.com

www.scribd.com/doc

www.soas.ac.uk

www.web-source.net

Notes

Unit 13: Test of Significance

CONTENTS

Objectives
Introduction
13.1 Small Sample Tests
13.1.1 T-test
13.1.2 Snedecor's F-distribution
13.2 Large Sample Test
13.2.1 Z-test (Parametric Test)
13.2.2 Chi-square Test
13.2.3 ANOVA
13.3 Summary
13.4 Keywords
13.5 Review Questions
13.6 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss the small sample tests;
- Explain the large sample test.

Introduction

Tests for statistical significance are used to estimate the probability that a relationship observed in the data occurred only by chance; the probability that the variables are really unrelated in the population. They can be used to filter out unpromising hypotheses. In research reports, tests of statistical significance are reported in three ways. First, the results of the test may be reported in the textual discussion of the results. Include:

Tanima Dutta, Lovely Professional University

1. Hypothesis
2. Test statistic used and its value
3. Degrees of freedom
4. Value for alpha (p-value)

Tests for statistical significance are used because they constitute a common yardstick that can be understood by a great many people, and they communicate essential information about a research project that can be compared to the findings of other projects. However, they do not assure that the research has been carefully designed and executed. In fact, tests for statistical significance may be misleading, because they are precise numbers. But they have no relationship to the practical significance of the findings of the research.

Finally, one must always use measures of association along with tests for statistical significance. The latter estimate the probability that the relationship exists; while the former estimate the

Notes

strength (and sometimes the direction) of the relationship. Each has its use, and they are best when used together.

There are two types of tests:

- Small Sample Tests
- Large Sample Test

13.1 Small Sample Tests

13.1.1 T-test

T-test is used in the following circumstances: When the sample size n<30.

Example: A certain pesticide is packed into bags by a machine. Random samples of 10 bags are drawn and their contents are found as follows: 50,49,52,44,45,48,46,45,49,45. Confirm whether the average packaging can be taken to be 50 kgs.

In this text, the sample size is less than 30. Standard deviations are not known using this test. We can find out if there is any significant difference between the two means i.e. whether the two population means are equal.

The Student's T-distribution

Let $X_1, X_2 \ldots\ldots X_n$ be n independent random variables from a normal population with mean *m* and standard deviation *s* (unknown).

When *s* is not known, it is estimated by s, the sample standard deviation $\left(s = \sqrt{\frac{1}{n-1}\sum\left(X_i - \bar{X}\right)^2}\right)$.

In such a case we would like to know the exact distribution of the statistic $\frac{\bar{X} - \mu}{s/\sqrt{n}}$ and the answer to this is provided by t-distribution.

W.S. Gosset defined t statistic as $t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ which follows t - distribution with (n–1) degrees of freedom.

Features of t-distribution

1. Like c^2- distribution, t-distribution also has one parameter *n* = n–1, where n denotes sample size. Hence, this distribution is known if *n* is known.
2. Mean of the random variable t is zero and standard deviation is $\sqrt{\frac{\nu}{\nu - 2}}$, for *n* > 2.
3. The probability curve of t-distribution is symmetrical about the ordinate at t = 0. Like a normal variable, the t variable can take any value from–∞ to ∞.
4. The distribution approaches normal distribution as the number of degrees of freedom become large.

Notes

5. The random variate t is defined as the ratio of a standard normal variate to the square root of χ^2 - variate divided by its degrees of freedom.

To show this we can write $t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{(\bar{X} - \mu)\sqrt{n}}{s}$

Dividing numerator and denominator by σ, we get

$$t = \frac{\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}}{\frac{s}{\sigma}} = \frac{\frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}}}{\sqrt{s^2/\sigma^2}} = \frac{\frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1} \cdot \frac{\sum(X_i - \bar{X})^2}{\sigma^2}}}$$

$$= \frac{\frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}}}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} = \frac{\text{Standard Normal Variate}}{\sqrt{\chi^2\text{-variate}}}$$

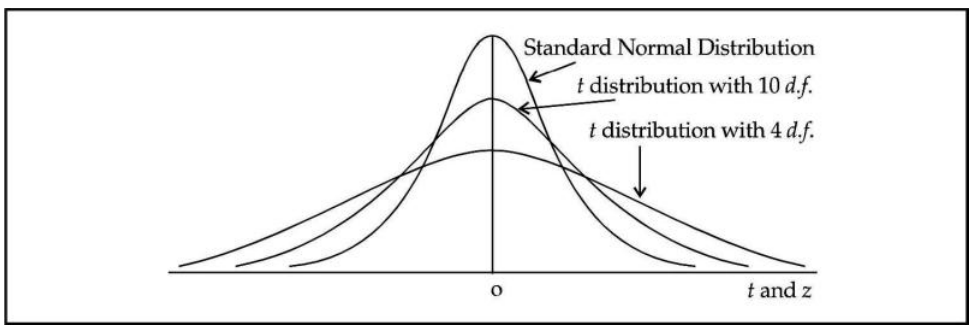


Illustration: There are two nourishment programmes 'A' and 'B'. Two groups of children are subjected to this. Their weight is measured after six months. The first group of children subjected to the programme 'A' weighed 44,37,48,60,41 kgs. at the end of programme. The second group of children were subjected to nourishment programme 'B' and their weight was 42, 42, 58, 64, 64, 67, 62 kgs. at the end of the programme. From the above, can we conclude that nourishment programme 'B' increased the weight of the children significantly, given a 5% level of confidence.

Null Hypothesis: There is no significant difference between Nourishment programme 'A' and 'B'.

Alternative Hypothesis: Nourishment programme B is better than 'A' or Nourishment programme 'B' increase the children's weight significantly.

Solution:

	Nourishment programme A			**Nourishment programme B**	
X	$x - \bar{x}$ = (x-46)	$(x - \bar{x})^2$	y	$y - \bar{y}$ =(y-57)	$(y - \bar{y})^2$
44	-2	4	42	-15	225
37	-9	81	42	-15	225

Contd...

Notes

48	2	4	58	1	1
60	14	196	64	7	49
41	-5	25	64	7	49
			67	10	100
			62	5	25
230	**0**	**310**	**399**	**0**	**674**

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Here $n_1 = 5$ $\quad n_1 = 7$

$\sum x = 230$ $\quad \sum y = 399$

$\sum(x - \bar{x})^2 = 310$ $\quad \sum(y - \bar{y})^2 = 674$

$$\bar{x} = \frac{\sum x}{n_1} = \frac{230}{5} = 46$$

$$\bar{y} = \frac{\sum y}{n_2} = \frac{399}{7} = 57$$

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left\{\sum(x - \bar{x})^2 + \sum(y - \bar{y})^2\right\}$$

D.F = $(n_1+n_2-2) = (5 + 7-2) = 10$

$$s^2 = \frac{1}{10}\{310 + 674\} = 98.4$$

$$t = \frac{46 - 57}{\sqrt{98.4 \times \left(\frac{1}{5} + \frac{1}{7}\right)}}$$

$$= \frac{-11}{\sqrt{98.4 \times \left(\frac{12}{35}\right)}}$$

$$= \frac{-11}{\sqrt{33.73}} = -\frac{11}{5.8}$$

$$= 1.89$$

t at 10 d.f. at 5% level is 1.81.

Since, calculated t is greater than 1.81, it is significant. Hence HA is accepted. Therefore the two nutrition programmes differ significantly with respect to weight increase.

Notes

13.1.2 Snedecor's F-distribution

Let there be two independent random samples of sizes n1 and n2 from two normal populations with variances σ_1^2 and σ_2^2 respectively. Further, let $s_1^2 = \frac{1}{n_1 - 1}\sum\left(X_{1i} - \bar{X}_1\right)^2$ and $s_2^2 = \frac{1}{n_2 - 1}\sum\left(X_{2i} - \bar{X}_2\right)^2$ be the variances of the first sample and the second samples respectively. Then F - statistic is defined as the ratio of two c^2 - variates. Thus, we can write

$$F = \frac{\frac{\chi^2_{n_1-1}}{n_1 - 1}}{\frac{\chi^2_{n_2-1}}{n_2 - 1}} = \frac{\frac{(n_1 - 1)s_1^2}{\sigma_1^2}/(n_1 - 1)}{\frac{(n_2 - 1)s_2^2}{\sigma_2^2}/(n_2 - 1)} = \frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}}$$

Features of F-distribution

1. This distribution has two parameters n_1 (= n_1 - 1) and n_2 (= n_2 - 1).

2. The mean of F - variate with n_1 and n_2 degrees of freedom is $\frac{\nu_2}{\nu_2 - 2}$ and standard error is $\left(\frac{\nu_2}{\nu_2 - 2}\right)\sqrt{\frac{2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)}}$.

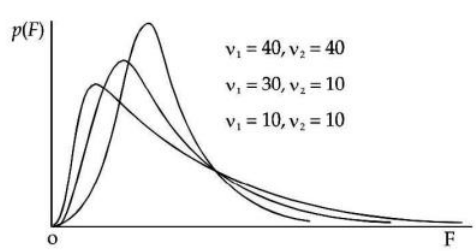
> *Notes* We note that the mean will exist if $\nu_2 > 2$ and standard error will exist if $\nu_2 > 4$. Further, the mean > 1.

3. The random variate F can take only positive values from 0 to ∞.

4. For large values of ν_1 and ν_2, the distribution approaches normal distribution.

5. If a random variate follows t-distribution with n degrees of freedom, then its square follows F-distribution with 1 and n d.f. i.e. $t_2^n = F_{1,\nu}$

6. F and χ^2 are also related as $F_{\nu 1'\nu 2} = \frac{(\chi^2_{\nu_1})}{\nu_1}$ as $\nu_2 \to \infty$



Notes

13.2 Large Sample Test

13.2.1 Z-test (Parametric Test)

(a) When sample size is > 30

P_1 = Proportion in sample 1

P_2 = Proportion in sample 2

Example: You are working as a purchase manager for a company. The following information has been supplied by two scooter tyres manufacturers.

	Company A	Company B
Mean life (in km)	13000	12000
S.D (in km)	340	388
Sample size	100	100

In the above, the sample size is 100, hence a Z-test may be used.

(b) Testing the hypothesis about difference between two means: This can be used when two population means are given and null hypothesis is Ho : P1 = P2.

Example: In a city during the year 2000, 20% of households indicated that they read 'Femina' magazine. Three years later, the publisher had reasons to believe that circulation has gone up. A survey was conducted to confirm this. A sample of 1,000 respondents were contacted and it was found 210 respondents confirmed that they subscribe to the periodical 'Femina'. From the above, can we conclude that there is a significant increase in the circulation of 'Femina'?

Solution:

We will set up null hypothesis and alternate hypothesis as follows:

Null Hypothesis is H_0. $\mu = 15\%$

Alternate Hypothesis is HA. $\mu > 15\%$

This is a one-tailed (right) test.

$$Z = \frac{P-\mu}{\sqrt{\frac{\mu(1-\mu)}{n}}}$$

$$Z = \frac{\frac{210}{1000}-0.20}{\sqrt{\frac{0.20(1-0.20)}{1000}}}$$

$$Z = \frac{0.21-0.20}{\sqrt{\frac{0.2\times 0.8}{1000}}}$$

$$= \frac{0.01-\mu}{\sqrt{\frac{0.16}{1000}}}$$

Notes

$$= \frac{0.1}{\frac{0.4}{31.62}}$$

$$= \frac{0.1}{0.012} = 8.33$$

As the value of Z at 0.05 =1.64 and calculated value of Z falls in the rejection region, we reject null hypothesis, and therefore we conclude that the sale of '*Femina*' has increased significantly.

13.2.2 Chi-square Test

With the help of this test, we will come to know whether two or more attributes are associated or not. How much the two attributes are related cannot be by Chi-Square test. Suppose, we have certain number of observations classified according to two attributes. We may like to know whether a newly introduced medicine is effective in the treatment of certain disease or not.

Caution One case where the distribution of the test statistic is an exact chi-square distribution is the test that the variance of a normally-distributed population has a given value based on a sample variance. Such a test is uncommon in practice because values of variances to test against are seldom known exactly.

The numbers of automobile accidents per week in a certain city were as follows:

Months	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct
No. of accidents	12	8	20	2	14	10	15	6	9	4

Does the above data indicate that accident conditions were uniform during the 10- month period.

Expected frequency $= 12 + 8 + 20 + 2 + 14 + 10 + 15 + 6 + 9 + 4 = \frac{100}{10} = 10$

Computation

Null hypothesis: The accident occurrence is uniform over a 10-week period.

Month	Observed No. of accidents	Expected No. of accidents	O – E	$(O - E)^2$	$\frac{(O-E)^2}{E}$
1	12	10	2	4	0.4
2	8	10	-2	4	0.4
3	20	10	10	100	10.0
4	2	10	-8	64	6.4
5	14	10	4	16	1.6
6	10	10	0	0	0.0
7	15	10	5	25	2.5
8	6	10	-4	16	1.6
9	9	10	-1	1	0.1
10	4	10	-6	36	3.6
	100	100	0		26.6

Notes

$$\therefore \qquad x^2 = \sum \frac{(O-E)^2}{E} = (26.6)$$

Where O is the observed frequency, E is the expected frequency.

$$D.F = 10 - 1 = 9$$

Table value at 5% for 9 degree of freedom = 16.91

Since calculated value = 26.6 greater than table value of 19.19, null hypothesis rejected at 5% level of significance.

Conclusion: The accident occurring are not uniform over a 10-week period.

Task What hypothesis, test and procedure would you use in the following situation?

1. An automobile company has manufacturing facility at two different geographical locations. Each location manufactures two-wheelers of a different model. The customer wants to know if the mileage given by both the models is the same or not. Samples of 45 numbers may be taken for this purpose.
2. A company has 22 sales executives. They underwent a training programme. The test must evaluate whether the sales performance is unchanged or improved after the training programme.
3. A company has three categories of managers:
   a. With professional qualifications but without work experience.
   b. With professional qualifications accompanied by work experience.
   c. Without professional qualifications but with work experience.

Self Assessment

Fill in the blanks:

1. There are two types of tests: Small Sample Tests and …….
2. With the help of ……. test, we will come to know whether two or more attributes are associated or not.

13.2.3 ANOVA

(a) ***ANOVA:*** It is a statistical technique. It is used to test the equality of three or more sample means. Based on the means, inference is drawn whether samples belongs to same population or not.

(b) ***Conditions for using ANOVA:***
   (1) Data should be quantitative in nature.
   (2) Data normally distributed.
   (3) Samples drawn from a population follows random variation.

(c) ***ANOVA can be discussed in two parts:***
   (1) One-way classification
   (2) Two and three-way classification.

One-way ANOVA

Notes

Following are the steps followed in ANOVA:

(a) Calculate the variance between samples.

(b) Calculate the variance within samples.

(c) Calculate F ratio using the formula.

F= Variance between the samples / Variance within the sample

(d) Compare the value of F obtained above in (c) with the critical value of F such as 5% level of significance for the applicable degree of freedom.

(e) When the calculated value of F is less than the table value of F, the difference in sample means is not significant and a null hypothesis is accepted. On the other hand, when the calculated value of F is more than the critical value of F, the difference in sample means is considered as significant and the null hypothesis is rejected.

Example: ANOVA is useful.

(1) To compare the mileage achieved by different brands of automotive fuel.

(2) Compare the first year earnings of graduates of half a dozen top business schools.

Application in Market Research

Consider the following pricing experiment. Three prices are considered for a new toffee box introduced by Nutrine company. Price of three varieties of toffee boxes are ₹ 39, ₹ 44 and ₹ 49. The idea is to determine the influence of price levels on sales. Five supermarkets are selected to exhibit these toffee boxes. The sales are as follows:

Price (₹)	1	2	3	4	5	Total	Sample mean $\bar{x}$
39	8	12	10	9	11	50	10
44	7	10	6	8	9	40	8
49	4	8	7	9	7	35	7

What the manufacturer wants to know is: (1) Whether the difference among the means is significant? If the difference is not significant, then the sale must be due to chance. (2) Do the means differ? (3) Can we conclude that the three samples are drawn from the same population or not?

Two-way ANOVA

The procedure to be followed to calculate variance is the same as it is for the one-way classification. The example of two-way classification of ANOVA is as follows:

Example: A firm has four types of machines – A , B, C and D. It has put four of its workers on each machines for a specified period, say one week. At the end of one week, the average output of each worker on each type of machine was calculated. These data are given below:

	Average production by the type of machine			
	A	B	C	D
Worker 1	25	26	23	28
Worker 2	23	22	24	27

Contd...

Notes

Worker 3	27	30	26	32
Worker 4	29	34	27	33

The firm is interested in knowing:

(a) Whether the mean productivity of workers is significantly different.

(b) Whether there is a significant difference in the mean productivity of different types of machines.

Illustration: Company 'X' wants its employees to undergo three different types of training programme with a view to obtain improved productivity from them. After the completion of the training programme, 16 new employees are assigned at random to three training methods and the production performance was recorded.

The training manager's problem is to find out if there are any differences in the effectiveness of the training methods? The data recorded is as under:

Daily output of new employees

Method 1	15	18	19	22	11	
Method 2	22	27	18	21	17	
Method 3	18	24	19	16	22	15

Following steps are followed.

1. Calculate Sample mean i.e. $\bar{x}$
2. Calculate General mean i.e. $\bar{x}$
3. Calculate variance between columns using the formula $\bar{\sigma}^2 = \frac{\sum n_i\left(x_i - \bar{\bar{x}}\right)^2}{k-1}$

   where K = (n_1 + n_2 + n_3-3).
4. Calculate sample variance. It is calculated using formula:

   Sample variance $s_i^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}$ where n is No. of observation under each method.
5. Calculate variance within columns using the formula $\bar{\sigma}^2 = \frac{\sum n_i - 1}{n_r - k}$
6. Calculate F using the ratio F = $\left(\frac{\text{between column variance}}{\text{within column variance}}\right)$
7. Calculate the number of degree of freedom in the numerator F ratio using equation, d.f = (No. of samples -1).
8. Calculate the number of degree of freedom in the denominator of F ratio using the equation d.f = $\sum\left(n_i - k\right)$
9. Refer to F table f8 find value.
10. Draw conclusions.

Solution:

Notes

Method 1	Method 2	Method 3
15	22	24
18	27	19
19	18	16
22	21	22
11	17	15
		18
85	**105**	**114**

1. Sample mean is calculated as follows:

$$\bar{x}_1 = \frac{85}{5} = 17 \quad \bar{x}_2 = \frac{105}{5} = 21 \quad \bar{x}_3 = \frac{114}{6} = 19$$

2. Grand mean

$$= \bar{\bar{x}} = \frac{15+18+19+22+11+22+27+18+21+17+24+19+16+22+15+18}{16}$$

$$= \frac{304}{16} = 19$$

3. Calculate variance between columns:

n	$\bar{x}$	$\bar{\bar{x}}$	$\bar{x}-\bar{\bar{x}}$	$\left(\bar{x}-\bar{\bar{x}}\right)^2$	$n\left(\bar{x}-\bar{\bar{x}}\right)^2$
5	17	19	-2	4	5 × 4 = 20
5	21	19	2	4	5 × 4 = 20
6	19	19	0	0	6 × 0 = 0
				$\sum n_i\left(\bar{x}_1-\bar{\bar{x}}\right)^2$	= 40

$$\bar{\sigma}^2 = \frac{\sum n_i\left(x_i-\bar{\bar{x}}\right)^2}{k-1} = \frac{40}{3-1} = 20$$

Variance between column = 20

4. Calculation sample variance:

Training method -1		Training method -2	Training method -3		
$x-\bar{x}$	$\left(x-\bar{x}\right)^2$	$x-\bar{x}$	$\left(x-\bar{x}\right)^2$	$x-\bar{x}$	
15-17	$(-2)^2 = 4$	22-21	$(1)^2 = 1$	18-19	$(1)^2 = 1$
18-17	$(1)^2 = 1$	27-21	$(6)^2 = 36$	24-19	$(5)^2 = 25$
19-17	$(2)^2 = 4$	18-21	$(-3)^2 = 9$	19-19	$(0)^2 = 0$
22-17	$(5)^2 = 25$	21-21	$(0)^2 = 1$	16-19	$(-3)^2 = 9$
11-17	$(-6)^2 = 36$	17-21	$(-4)^2 = 16$	22-19	$(3)^2 = 9$
				15-19	$(-4)^2 = 16$
$\sum\left(x-\bar{x}\right)^2 = 70$		$\sum\left(x-\bar{x}\right)^2 = 62$			$\sum\left(x-\bar{x}\right)^2 = 60$

Notes

Sample variance = $\frac{\sum(x-\bar{x})^2}{n-1} = \frac{70}{5-1}$, $\frac{\sum(x-\bar{x})^2}{n-1} = \frac{62}{5-1}$, $\frac{\sum(x-\bar{x})^2}{n-1} = \frac{60}{6-1}$

$s_1^2 = \frac{70}{4} = 17.5 \qquad s_2^2 = \frac{62}{4} = 15.5 \qquad s_3^2 = \frac{60}{5} = 12$

5. Within column variance $\bar{\sigma}^2 = \sum\left(\frac{n_i - 1}{n_i - k}\right)s_1^2$

$$= \left(\frac{5-1}{16-3}\right)\times 17.5 + \left(\frac{5-1}{16-3}\right)\times 15.5 + \left(\frac{6-1}{16-3}\right)\times 12$$

$$= \left(\frac{4}{13}\right)\times 17.5 + \left(\frac{4}{13}\right)\times 15.5 + \frac{5}{13}\times 12$$

Within column variance $= \frac{192}{13} = 14.76$

6. F = $\frac{\text{Between column variance}}{\text{Within column variance}} = \frac{20}{14.76} = 1.354$

7. d.f of Numerator = (3 – 1) = 2.

8. d.f of Denominator = $\sum n_1 - k$ = (5 – 1) + (5 – 1) + (6 – 1) = 16 – 3 = 13.

9. Refer to table using d.f = 2 and d.f = 13.

10. The value is 3.81. This is the upper limit of acceptance region. Since calculated value 1.354 lies within it we can accept H0, the null hypothesis.

Conclusion: There is no significant difference in the effect of the three training methods.

Self Assessment

Fill in the blanks:

3. For using ANOVA, the data should be ..... in nature.

4. F test has ..... parameters.

13.3 Summary

- Testing the hypothesis about difference between two means: This can be used when two population means are given and null hypothesis is H_o : P1 = P2.
- ***ANOVA*** is a statistical technique. It is used to test the equality of three or more sample means. Based on the means, inference is drawn whether samples belongs to same population or not.

13.4 Keywords

ANOVA: It is a statistical technique. It is used to test the equality of three or more sample means. Based on the means, inference is drawn whether samples belongs to same population or not.

Notes

Significance Level: Significance level is the criterion used for rejecting the null hypothesis.

Tests for statistical significance: Tests for statistical significance are used to estimate the probability that a relationship observed in the data occurred only by chance; the probability that the variables are really unrelated in the population.

13.5 Review Questions

1. Each person in a random sample of 50 was asked to state his/her sex and preferred colour. The resulting frequencies are shown below.

Colour		Red	Blue	Green
	Male	5	14	6
Sex	**Female**	15	6	4

A chi-square test is used to test the null hypothesis that sex and preferred colour are independent. Will you reject at the null hypothesis 0.005 level? Why/ Why not?

2. Are all employees equally prone to having accidents? To investigate this hypothesis, Parry (1985) looked at a light manufacturing plant and classified the accidents by type and by age of the employee.

Age	Accident Type		
	Sprain	**Burn**	**Cut**
Under 25	9	17	5
25 or over	61	13	12

A chi-square test gave a test-statistic of 20.78. If we test at a =.05, does the proportion of sprain, cuts and burns seems to be similar for both age classes? Why/ why not?

3. In hypothesis testing, if is the probability of committing an error of Type II. The power of the test, is then the probability of rejecting H0 when HA is true or not? Why?

4. In a statistical test of hypothesis, what would happen to the rejection region if , the level of significance, is reduced?

5. During the pre-flight check, Pilot Mohan discovers a minor problem - a warning light indicates that the fuel gauge may be broken. If Mohan decides to check the fuel level by hand, it will delay the flight by 45 minutes. If he decides to ignore the warning, the aircraft may run out of fuel before it gets to Mumbai. In this situation, what would be:

   (a) the appropriate null hypothesis? and;

   (b) a type I error?

Answers: Self Assessment

1. Large Sample Test
2. Chi-square
3. Quantitative
4. 2

Notes

13.6 Further Readings

Books

Abrams, M.A, *Social Surveys and Social Action*, London: Heinemann, 1951.

Arthur, Maurice, *Philosophy of Scientific Investigation* , Baltimore: John Hopkins University Press, 1943.

R.S. Bhardwaj, *Business Statistics*, Excel Books, New Delhi, 2008.

S.N. Murthy and U. Bhojanna, *Business Research Methods*, Excel Books, 2007.

Online links

www.indiastudychannel.com

www.scribd.com/doc

www.soas.ac.uk

www.web-source.net

Notes

Unit 14: Multivariate Analysis

CONTENTS

Objectives

Introduction

14.1 Discriminant Analysis

14.2 Factor Analysis

14.3 Cluster Analysis

14.3.1 Cluster Analysis on Three Dimensions

14.4 Conjoint Analysis

14.5 Multidimensional Scaling (MDS)

14.5.1 Types of MDS

14.6 Summary

14.7 Keywords

14.8 Review Questions

14.9 Further Readings

Objectives

After studying this unit, you will be able to:

- Explain the multiple regressions;
- Discuss the discriminant analysis and conjoint analysis;
- Explain the factor analysis and cluster analysis;
- Describe the Multidimensional Scaling (MDS).

Introduction

As the name indicates, multivariate analysis comprises a set of techniques dedicated to the analysis of data sets with more than one variable. Several of these techniques were developed recently in part because they require the computational capabilities of modern computers. Multivariate analysis (MVA) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest. Sometimes, the marketers will come across situations, which are complex involving two or more variables. Hence, bi-variate analysis deals with this type of situation. Chi-Square is an example of bi-variate analysis. In multi-variate analysis, the numbers of variables to be tackled are many.

Example: The demand for television sets may depend not only on price, but also on the income of households, advertising expenditure incurred by TV manufacturer and other similar factors. To solve this type of problem, multivariate analysis is required.

Notes

Notes ***Multiple-variate analysis:*** This can be studied under:

1. Discriminant analysis
2. Factor analysis
3. Cluster analysis
4. Conjoint analysis
5. Multidimensional scaling.

14.1 Discriminant Analysis

In this analysis, two or more groups are compared. In the final analysis, we need to find out whether the groups differ one from another.

Example: Where discriminant analysis is used

1. Those who buy our brand and those who buy competitors' brand.
2. Good salesman, poor salesman, medium salesman.
3. Those who go to Food World to buy and those who buy in a Kirana shop.
4. Heavy user, medium user and light user of the product.

Suppose there is a comparison between the groups mentioned as above along with demographic and socio-economic factors, then discriminant analysis can be used. One way of doing this is to proceed and calculate the income, age, educational level, so that the profile of each group could be determined. Comparing the two groups based on one variable alone would be informative but it would not indicate the relative importance of each variable in distinguishing the groups. This is because several variables within the group will have some correlation which means that one variable is not independent of the other.

If we are interested in segmenting the market using income and education, we would be interested in the total effect of two variables in combinations, and not their effects separately. Further, we would be interested in determining which of the variables are more important or had a greater impact. To summarize, we can say, that Discriminant Analysis can be used when we want to consider the variables simultaneously to take into account their interrelationship.

Like regression, the value of dependent variable is calculated by using the data of independent variable.

$Z = b_1x_1 + b_2x + b_3x_3 + ...$

Z = Discriminant score

b_1 = Discriminant weight for variable

x = Independent variable

As can be seen in the above, each independent variable is multiplied by its corresponding weightage.

This results in a single composite discriminant score for each individual. By taking the average of discriminant score of the individuals within a certain group, we create a group mean. This is known as centroid. If the analysis involves two groups, there are two centroids. This is very similar to multiple regression, except that different types of variables are involved.

Application: A company manufacturing FMCG products introduces a sales contest among its marketing executives to find out "How many distributors can be roped in to handle the company's product". Assume that this contest runs for three months. Each marketing executive is given target regarding number of new distributors and sales they can generate during the period. This

Notes

target is fixed and based on the past sales achieved by them about which, the data is available in the company. It is also announced that marketing executives who add 15 or more distributors will be given a Maruti omni-van as prize. Those who generate between 5 and 10 distributors will be given a two-wheeler as the prize. Those who generate less than 5 distributors will get nothing. Now assume that 5 marketing executives won a Maruti van and 4 won a two-wheeler.

The company now wants to find out, "Which activities of the marketing executive made the difference in terms of winning a prize and not winning the prize". One can proceed in a number of ways. The company could compare those who won the Maruti van against the others. Alternatively, the company might compare those who won, one of the two prizes against those who won nothing. It might compare each group against each of the other two.

Discriminant analysis will highlight the difference in activities performed by**each group members** to get the prize. The activity might include:

1. More number of calls made to the distributors.
2. More personal visits to the distributors with advance appointments.
3. Use of better convincing skills.

Discriminant Analysis

1. What variable discriminates various groups as above; the number of groups could be two or more. Dealing with more than two groups is called Multiple Discriminant Analysis (M.D.A).
2. Can discriminating variables be chosen to forecast the group to which the brand/person/place belong to?
3. Is it possible to estimate the size of different groups?

Self Assessment

Fill in the blanks:

1. ..... analysis is used if there are more than 2 variables.
2. An advantage of the non-metric models is that they permit the researcher to ..... and ..... preference data.
3. ..... analysis is used by those who buy our brand and those who buy competitors' brand.

14. 2 Factor Analysis

The main purpose of Factor Analysis is to group large set of variable factors into fewer factors. Each factor will account for one or more component. Each factor a combination of many variables. There are two most commonly employed factor analysis procedures. They are:

(1) Principle component analysis

(2) Common factor analysis.

When the objective is to summarise information from a large set of variables into fewer factors, principle component factor analysis is used. On the other hand, if the researcher wants to analyse the components of the main factor, common factor analysis is used.

Example: **Common factor –** Inconvenience inside a car. The components may be:

1. Leg room.
2. Seat arrangement.

Notes

3. Entering the rare seat.
4. Inadequate dickey space.
5. Door locking mechanism.

Principle Component Factor Analysis

Purposes: Customer feedback about a two-wheeler manufactured by a company.

Method: The M.R manager prepares a questionnaire to study the customer feedback. The researcher has identified six variables or factors for this purpose. They are as follows:

1. Fuel efficiency (A)
2. Durability (Life) (B)
3. Comfort (C)
4. Spare parts availability (D)
5. Breakdown frequency (E)
6. Price (F)

The questionnaire may be administered to 5,000 respondents. The opinion of the customer is gathered. Let us allot points 1 to 10 for the variables factors A to F. 1 is the lowest and 10 is the highest. Let us assume that application of factor analysis has led to grouping the variables as follows:

A, B, D, E into factor – 1

F into Factor -2

C into Factor - 3

Factor - 1 can be termed as Technical factor;

Factor - 2 can be termed as Price factor;

Factor - 3 can be termed as Personal factor.

For future analysis, while conducting a study to obtain customers' opinion, three factors mentioned above would be sufficient. One basic purpose of using factor analysis is to reduce the number of independent variables in the study. By having too many independent variables, the M.R study will suffer from following disadvantages:

1. Time for data collection is very high due to several independent variables.
2. Expenditure increases due to the time factor.
3. Computation time is more, resulting in delay.
4. There may be redundant independent variables.

Did u know? **What is correspondence analysis?**

Correspondence analysis is a descriptive/exploratory technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns. The results provide information which is similar in nature to those produced by Factor Analysis techniques, and they allow one to explore the structure of categorical variables included in the table. The most common kind of table of this type is the two-way frequency cross-tabulation table.

In a typical correspondence analysis, a cross-tabulation table of frequencies is first standardized, so that the relative frequencies across all cells sum to 1.0. One way to state the goal of a typical analysis is to represent the entries in the table of relative frequencies in terms of the distances between individual rows and/or columns in a low-dimensional space.

Self Assessment

Notes

Fill in the blanks:

4. In factor analysis, the expenditure ..... due to time factor.
5. Correspondence analysis is a ..... technique.

14.3 Cluster Analysis

Cluster Analysis is used:

1. To classify persons or objects into small number of clusters or group.
2. To identify specific customer segment for the company's brand.

Cluster Analysis is a technique used for classifying objects into groups. This can be used to sort data (a number of people, companies, cities, brands or any other objects) into homogeneous groups based on their characteristics.

The result of Cluster Analysis is a grouping of the data into groups called clusters. The researcher can analyse the clusters for their characteristics and give the cluster, names based on these.

Where can Cluster Analysis be applied?

The marketing application of cluster analysis is in customer segmentation and estimation of segment sizes. Industries, where this technique is useful include automobiles, retail stores, insurance, B-to-B, durables and packaged goods. Some of the well-known frameworks in consumer behaviour (like VALS) are based on value cluster analysis.

Cluster Analysis is applicable when:

- An FMCG company wants to map the profile of its target audience in terms of lifestyle, attitude and perceptions.
- A consumer durable company wants to know the features and services a consumer takes into account, when purchasing through catalogues.
- A housing finance corporation wants to identify and cluster the basic characteristics, lifestyles and mindset of persons who would be availing housing loans. Clustering can be done based on parameters such as interest rates, documentation, processing fee, number of installments etc.

Process

There are two ways in which Cluster Analysis can be carried out:

1. First, objects/respondents are segmented into a pre-decided number of clusters. In this case, a method called non-hierarchical method can be used, which partitions data into the specified number of clusters
2. The second method is called the hierarchical method.

The above two are basic approaches used in cluster analysis. This can be used to segment customer groups for a brand or product category, or to segment retail stores into similar groups based on selected variables.

Interpretation of Results

Ideally, the variables should be measured on an interval or ratio scale. This is because the clustering techniques use the distance measure to find the closest objects to group into a cluster. An example of its use can be clustering of towns similar to each other which will help decide

Notes

where to locate new retail stores.

If clusters of customers are found based on their attitudes towards new products and interest in different kinds of activities, an estimate of the segment size for each segment of the population can be obtained, by looking at the number of objects in each cluster.

Names can also be given to clusters to describe each one. For example, there can be a cluster called "neo-rich". Segments are prioritised based on their estimated size.

Marketing strategies for each segment are fine-tuned based on the segment characteristics. For instance, a segment of customers, like sports car, get a special promotional offer during specific period.

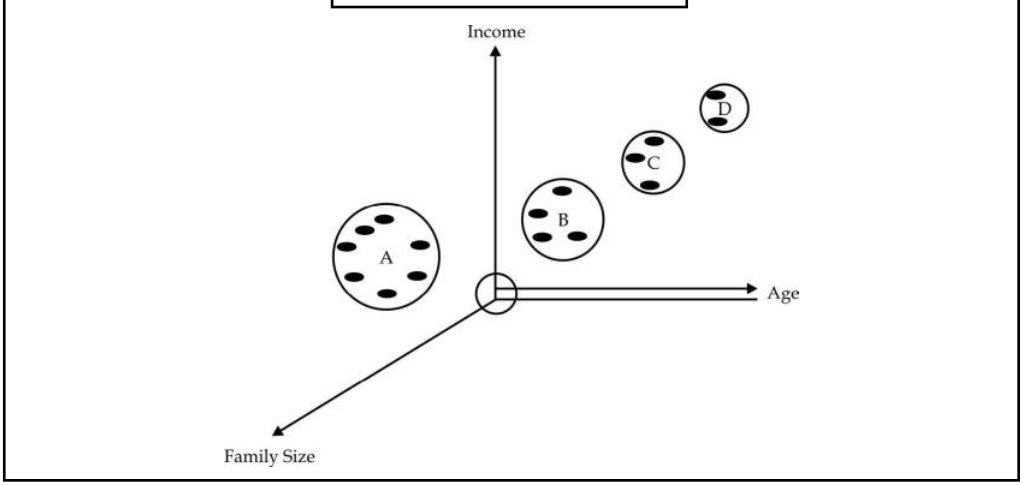
Example: In cluster analysis, the following five steps to be used:

1. Selection of the sample to be clustered (buyers, products, employees)
2. Definition on which the measurement to be made (Eg: product attributes, buyer characteristics, employees' qualification)
3. Computing the similarities among the entities.
4. Arrange the cluster in a hierarchy.
5. Cluster comparison and validation.

14.3.1 Cluster Analysis on Three Dimensions

The example below shows Cluster Analysis based on three dimensions age, income and family size. Cluster Analysis is used to segment the car-buying population in a Metro. For example "A" might represent potential buyers of low end cars. Example: Maruti 800 (for common man). These are people who are graduating from the two-wheeler market segment. Cluster "B" may represent mid-population segment buying Zen, Santro, Alto etc. Cluster "C" represents car buyers, who belong to upper strata of society. Buyers of Lancer, Honda city etc. Cluster "D" represents the super-rich cluster, i.e. Buyers of Benz, BMW etc.

Figure 14.1: Matching Measure



Example: Suppose there are five attributes, 1 to 5, on which we are judging two objects A and B. The existence of an attribute may be indicated by 1 and its absence by 0. In this way, two objects are viewed as similar if they share common attributes.

Notes

Table

Attribute	1	2	3	4	5	6	7
Brand – A	1	0	0	1	0	0	1
Brand – B	0	0	1	1	1	0	0

One measure of simple matching S is given by:

$$S = \frac{a+d}{a+b+c+d}$$

Where a = No. of attributes possessed by brands A and B

b = No. of attributes possessed by brand A but not by brand B

c = No. of attributes possessed by brand B but not by brand A

d = No. of attributes not possessed by both brands.

Substituting, we get $S = \frac{1+2}{1+2+2+2} = \frac{3}{7} = 0.43$

A and B's association is to be the extent of 43%.

It is now clear that object A possess attributes 1, 4, and 7 while object B possess the attributes 3, 4 and 5. A glance at the above table will indicate that objects A and B are similar in respect of 2 (0 & 0), 6 (0 & 0) and 4 (1 & 1). In respect of other attributes, there is no similarity between A and B. Now we can arrive at a simple matching measure by (a) counting up the total number of matches – either 0, 0 or 1, (b) dividing this number by the total number of attributes.

Symbolically SAB = M / N

SAB = Similarity between A and B

M = Number of attributes held in common (0 or 1)

N = Total number of attributes

SAB = 3 / 7 = 0.43

i.e., A & B are similar to the extent of 43%.

Self Assessment

Fill in the blanks:

6. In a typical correspondence analysis, a cross-tabulation table of frequencies is first .....
7. ..... Analysis is a technique used for classifying objects into groups.
8. The ..... application of cluster analysis is in customer segmentation and estimation of segment sizes.

14.4 Conjoint Analysis

Conjoint analysis is concerned with the measurement of the joint effect of two or more attributes that are important from the customers' point of view. In a situation where the company would like to know the most desirable attributes or their combination for a new product or service, the use of conjoint analysis is most appropriate.

Notes

Example: An airline would like to know, which is the most desirable combination of attributes to a frequent traveller: (a) Punctuality (b) Air fare (c) Quality of food served on the flight and (d) Hospitality and empathy shown.

Conjoint Analysis is a multivariate technique that captures the exact levels of utility that an individual customer places on various attributes of the product offering. Conjoint Analysis enables a direct comparison.

Example: A comparison between the utility of a price level of ₹ 400 versus ₹ 500, a delivery period of 1 week versus 2 weeks, or an after-sales response of 24 hours versus 48 hours.

Once we know the utility levels for each attribute (and at individual levels as well), we can combine these to find the best combination of attributes that gives the customer the highest utility, the second best combination that gives the second highest utility, and so on. This information is then used to design a product or service offering.

Application

Conjoint Analysis is extremely versatile and the range of applications includes virtually in any industry. New product or service design, including the concepts in the pre-prototyping stage can specifically benefit from the conjoint applications.

Some examples of other areas where this technique can be used are:

- Designing an automobile loan or insurance plan in the insurance industry,
- Designing a complex machine for business customers.

Process

Design attributes for a product are first identified. For a shirt manufacturer, these could be design such as designer shirts Vs plain shirts, this price of ₹ 400 versus ₹ 800. The outlets can have exclusive distribution or mass distribution. All possible combinations of these attribute levels are then listed out. Each design combination will be ranked by customers and used as input data for Conjoint Analysis. Then the utility of the products relative to price can be measured.

The output is a part-worth or utility for each level of each attribute. For *example,* the design may get a utility level of 5 and plain, 7.5. Similarly, the exclusive distribution may have a part utility of 2, and mass distribution, 5.8. We then put together the part utilities and come up with a total utility for any product combination we want to offer, and compare that with the maximum utility combination for this customer segment.

This process clarifies to the marketer about the product or service regarding the attributes that they should focus on in the design.

If a retail store finds that the height of a shelf is an important attribute for selling at a particular level, a well-designed shelf may result from this knowledge. Similarly, a designer of clocks will benefit from knowing the utility attached by customers to the dial size, background colours, and price range of the clocks.

Approach

From a discussion with the client, identify the design attributes to be studied and the levels at which they can be offered. Then build a list of product concepts on offer. These product concepts are then ranked by customers. Once this data is available, use Conjoint Analysis to derive the part utilities of each attribute level. This is then used to predict the best product design for the given customer segment. Use the **SPSS** Conjoint procedure to analyse the data.

Notes

There are three steps in **conjoint analysis:**

(a) Identification of relevant products or service attributes.

(b) Collection of data.

(c) Estimation of worth for the attribute chosen.

For attributes selection, the market researcher can conduct interview with the customers directly.

Example of conjoint analysis for a **Laptop:**

For a laptop, consider 3 **attributes:**

Weight (3 Kg or 5 Kg)

Battery life (2 hours or 4 hours)

Brand name (Lenovo or Dell)

Task Rank order the following combination of these characteristics:

1 = Most preferred, 8= Least preferred

Combination	Rank
3 Kg, 2 hours, Lenovo	4
5 Kg, 4 hours, Dell	5
5 Kg, 2 hours, Lenovo	8
3 Kg, 4 hours, Lenovo	3
3 Kg, 2 hours, Dell	2
5 Kg, 4 hours, Lenovo	7
5 Kg, 2 hours, Dell	6
3 Kg, 4 hours, Dell	1

One combination 3 kg, 4 hours, Dell clearly dominates and 5 kg, 2 hours, Lenovo is least preferred.

Let us now take the average rank for 3 kg option = 4+3+2+1 / 4 = 2.5

For 5 kg option average rank is 5+8+7+6 / 4 = 6.5

For 4 hour option 5+3+7+1 / 4 = 4

For 2 hour option 4+8+2+6 / 4 = 5

For Dell 5+6+1+2 / 4 = 3.5

For Lenovo 5.5

Looking at the difference in average ranks, the most important characteristic to this respondent is weight = 4, followed by brand name = 2 and battery life = 1.

Self Assessment

Fill in the blanks:

9. ..... analysis is concerned with the measurement of the joint effect of two or more attributes.

10. For ..... selection, the market researcher can conduct interview with the customers directly.

Notes

14.5 Multidimensional Scaling (MDS)

In addition to fulfilling the goals of detecting underlying structure and data reduction that is shares with other methods, multidimensional scaling (MDS) provides the researcher with a spatial representation of data that can facilitate interpretation and reveal relationships. Therefore, we can define MDS as "a set of multivariate statistical methods for estimating the parameters in and assessing the fit of various spatial distance models for proximity data."

The spatial display of data provided by MDS is why it is also sometimes referred to as perceptual mapping. MDS has much more flexibility about the types of data that can be used to generate the solution. Almost any measures of similarity and dissimilarity can be used, depending on what your statistical computer software will accept.

14.5.1 Types of MDS

In general, there are two types of MDS:

1. Metric
2. Non-metric

Metric MDS makes the assumption that the input data is either ratio or interval data, while the non-metric model requires simply that the data be in the form of ranks. Therefore, the non-metric model has more fewer restrictions than the metric model, but also less rigor. One technique to use if you are unsure whether your data is ordinal or can be considered interval is to try both metric and non-metric models. If the results are very close, the metric model may be used.

An advantage of the non-metric models is that they permit the researcher to categorize and examine preference data, such as the kind obtained in marketing studies or other areas where comparisons are useful.

Another technique, correspondence analysis, can work with categorical data, i.e., data at the nominal level of measurement, however that technique will not be described here.

Similarities and Differences between Factor Analysis and MDS

We have already seen that MDS can accept more different measures of similarity and dissimilarity than factor analysis techniques can. In addition, there are some differences in terminology. These differences reflect the origin of MDS in the field of psychology. The measure corresponding to factors are called alternatively dimensions or stimulus coordinates.

The output of MDS looks very similar to that of factor analysis and the determination of the optimal number of dimensions is handled in much the same way.

Steps in using MDS

There are four basic steps in MDS:

1. Data collection and formation of the similarity/dissimilarity matrix
2. Extraction of stimulus coordinates
3. Decision about the number of stimulus coordinates that represent the data
4. Rotation and interpretation

Example: An example of MDS

Let us say that you have a matrix of distances between a number of major cities, such as you might find on the back of a road map. These distances can be used as the input data to derive an

MDS solution. When the results are mapped in two dimensions, the solution will reproduce a conventional map, except that the MDS plot might need to be rotated so that the north-south and east-west dimensions conform to expectations. However, the once the rotation is completed, the configuration of the cities will be spatially correct.

Task Which technique would you use to measure the joint effect of various attributes while designing an automobile loan and why?

Self Assessment

Fill in the blanks:

11. When the objective is to summarise information from a large set of variables into fewer factors, ... analysis is used.
12. The ..... is a part-worth or utility for each level of each attribute.

14.6 Summary

- Multivariate analysis is used if there are more than 2 variables.
- Some of the multi variate analysis are discriminant analysis, Factor analysis, Cluster analysis, conjoint analysis, and multi dimensional scaling.
- In discriminant analysis, it is verified whether the 2 groups differ from one another.
- Factor analysis is used to reduce large no of various factors into fewer variables cluster analysis is used to segmenting the market or to identify the target group.
- MDS as a set of multivariate statistical methods for estimating the parameters in and assessing the fit of various spatial distance models for proximity data.
- The output of MDS looks very similar to that of factor analysis and the determination of the optimal number of dimensions is handled in much the same way.

14.7 Keywords

Cluster analysis: Cluster Analysis is a technique used for classifying objects into groups.

Discriminant analysis: In this analysis, two or more groups are compared. In the final analysis, we need to find out whether the groups differ one from another.

Multivariate analysis: In multi variate analysis, the number of variables to be tackled are many.

14.8 Review Questions

1. Do you think that the conjoint analysis will be useful in any manner for an airline? If yes how, if no, give an example where you think the technique is of immense help.
2. In your opinion, what are the main advantages of cluster analysis?
3. Which analysis would you use in a situation when the objective is to summarise information from a large set of variables into fewer factors? What will be the steps you would follow?
4. Which analysis would answer if it is possible to estimate the size of different groups?
5. Which analysis would you use to compare a good, bad and a mediocre doctor and why?
6. Analyse the weakness of principle component factor analysis.

Notes

7. Which multivariate analysis would you apply to identify specific customer segment for a company's brand and why?
8. Critically evaluate multidimensional scaling.

Answers: Self Assessment

1. Multivariate
2. Categorize, examine
3. Discriminant
4. Increases
5. Descriptive/exploratory
6. Standardized
7. Cluster
8. Marketing
9. Conjoint
10. Attributes
11. Principle component factor
12. Output

14.9 Further Readings

Books

A Prasuraman, Dhruv Grewal, *Marketing Research*, Biztantra.

Alan T Shao, *Marketing Research*, Cengage.

Cisnal Peter, *Marketing Research*, MCGE.

GA Chrchil, *Marketing Research*, Iacobucci, Thomson.

GC Beri, *Marketing Research*, TMH.

Hague & Morgan, *Marketing Research in Practice*, Kogan page.

OR Krishna Swamy, *Methodology of Research in Social Sciences*, HPH.

Paneerselvam, R, *Research Methods*, PHI.

Tull and Donalds, *Marketing Research*, MMIL.

Online links

www.indiastudychannel.com

www.scribd.com/doc

www.soas.ac.uk

www.web-source.net